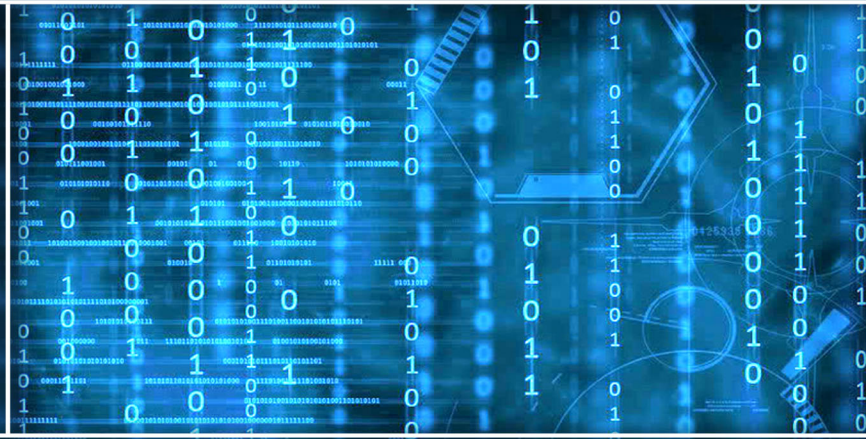


Volume 14 Issue 9

September 2023



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 9 September 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*

**Doroła 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: Data Anomaly Detection in the Internet of Things: A Review of Current Trends and Research Challenges

*Authors: Min Yang, Jiajie Zhang*

**PAGE 1 – 10**

Paper 2: Segmentation of Motion Objects in Video Frames using Deep Learning

*Authors: Feng JIANG, Jiao LIU, Jiya TIAN*

**PAGE 11 – 20**

Paper 3: Classification of Coherence Indices Extracted from EEG Signals of Mild and Severe Autism

*Authors: Lingyun Wu*

**PAGE 21 – 29**

Paper 4: Enhancing Breast Cancer Diagnosis using a Modified Elman Neural Network with Optimized Algorithm Integration

*Authors: Linkai Chen, CongZhe You, Honghui Fan, Hongjin Zhu*

**PAGE 30 – 38**

Paper 5: Stacked LSTM and Kernel-PCA-based Ensemble Learning for Cardiac Arrhythmia Classification

*Authors: Azween Abdullah, S. Nithya, M.Mary Shanthi Rani, S. Vijayalakshmi, Balamurugan Balusamy*

**PAGE 39 – 48**

Paper 6: A Versatile Shuffle Resource Units Recomputation Algorithm for Uplink OFDMA Random Access

*Authors: Azyyati Adiah Zazali, Shamala Subramaniam, Zuriati Ahmad Zukarnain, Abdullah Muhammed*

**PAGE 49 – 56**

Paper 7: The Promise of Self-Supervised Learning for Dental Caries

*Authors: Tran Quang Vinh, Haewon Byeon*

**PAGE 57 – 61**

Paper 8: Optimization Method for Trajectory Data Based on Satellite Doppler Velocimetry

*Authors: Junzhuo Li, Wenyong Li, Guan Lian*

**PAGE 62 – 69**

Paper 9: Optimized YOLOv7 for Small Target Detection in Aerial Images Captured by Drone

*Authors: Yanxin Liu, Shuai Chen, Lin Luo*

**PAGE 70 – 79**

Paper 10: DevOps Implementation Challenges in the Indonesian Public Health Organization

*Authors: Muhammad Yazid Al Qahar, Teguh Raharjo*

**PAGE 80 – 93**

Paper 11: A Bibliometric Analysis of Smart Home Acceptance by the Elderly (2004-2023)

*Authors: Bo Yuan, Norazlynn Kamal Basha*

**PAGE 94 – 104**

Paper 12: Automatic Generation of Image Caption Based on Semantic Relation using Deep Visual Attention Prediction

*Authors: M. M. EL-GAYAR*

**PAGE 105 – 114**

Paper 13: Enhancing Oil Price Forecasting Through an Intelligent Hybridized Approach

Authors: Hicham BOUSSATTA, Marouane CHIHAB, Younes CHIHAB, Mohammed CHINY

PAGE 115 – 125

Paper 14: Compression Analysis of Hybrid Model Based on Scalable WDR Method and CNN for ROI-based Medical Image Transmission

Authors: Bindulal T.S

PAGE 126 – 135

Paper 15: A Proposed Intelligent Model with Optimization Algorithm for Clustering Energy Consumption in Public Buildings

Authors: Ahmed Abdelaziz, Vitor Santos, Miguel Sales Dias

PAGE 136 – 152

Paper 16: A Survey of Evolving Performance Analysis Technologies, Algorithms and Models for Sports

Authors: Shamala Subramaniam, Manoj Ravi Shankar, Azyyati Adiah Zazali, Hong Siaw Swin, Zarina Muhamed, Sivakumar Rajagopal, Mohamad Zamri Napiah, Faisal Embung

PAGE 153 – 161

Paper 17: An Improvement for Spatial-Temporal Queries of ATMGRAPH

Authors: ZHANG Zhiyuan, HAN Boyang

PAGE 162 – 168

Paper 18: Comparison of Machine Learning Algorithms for Crime Prediction in Dubai

Authors: Shaikha Khamis AlAbdouli, Ahmad Falah Alomosh, Ali Bou Nassif, Qassim Nasir

PAGE 169 – 173

Paper 19: Preserving Cultural Heritage Through AI: Developing LeNet Architecture for Wayang Image Classification

Authors: Muhathir, Nurul Khairina, Rehia Karenina Isabella Barus, Mutammimul Ula, Ilham Sahputra

PAGE 174 – 181

Paper 20: A Comprehensive Review of Modern Methods to Improve Diabetes Self-Care Management Systems

Authors: Alhuseen Omar Alsayed, Nor Azman Ismail, Layla Hasan, Farhat Embarak

PAGE 182 – 203

Paper 21: An Improved Convolutional Neural Network for Churn Analysis

Authors: Priya Gopal, Nazri Bin MohdNawi

PAGE 204 – 210

Paper 22: A New Method for Classifying Intracerebral Hemorrhage (ICH) Based on Diffusion Weighted –Magnetic Resonance Imaging (DW-MRI)

Authors: Andi Kurniawan Nugroho, Jajang Edi Priyanto, Dinar Mutiara Kusumo Nugraheni

PAGE 211 – 217

Paper 23: Application Prototype for Inclusive Literacy for People with Reading Disabilities

Authors: Laberiano Andrade-Arenas, Roberto Santiago Bellido-García, Pedro Molina-Velarde, Cesar Yactayo-Arias

PAGE 218 – 225

Paper 24: LAD-YOLO: A Lightweight YOLOv5 Network for Surface Defect Detection on Aluminum Profiles

Authors: Dongxue Zhao, Shenbo Liu, Yuanhang Chen, Da Chen, Zhelun Hu, Lijun Tang

PAGE 226 – 234

**Paper 25: Improved YOLO-X Model for Tomato Disease Severity Detection using Field Dataset**

*Authors: Rajasree R, C Beulah Christalin Latha*

**PAGE 235 – 242**

**Paper 26: He and She in Video Games: Impact of Gender on Video Game Participation and Perspectives**

*Authors: Deena Alghamdi*

**PAGE 243 – 249**

**Paper 27: Hyperparameter Tuning of Semi-Supervised Learning for Indonesian Text Annotation**

*Authors: Siti Khomsah, Nur Heri Cahyana, Agus Sasmito Aribowo*

**PAGE 250 – 256**

**Paper 28: Usability Testing of Memorable Word in Security Enhancing in e-Government and e-Financial Systems**

*Authors: Hanan Alotaibi, Dania Aljeaid, Amal Alharbi*

**PAGE 257 – 265**

**Paper 29: Enhanced Brain Tumor Detection and Classification in MRI Scans using Convolutional Neural Networks**

*Authors: Ruqsar Zaitoon, Hussain Syed*

**PAGE 266 – 275**

**Paper 30: Method for Hyperparameter Tuning of Image Classification with PyCaret**

*Authors: Kohei Arai, Jin Shimazoe, Mariko Oda*

**PAGE 276 – 282**

**Paper 31: A Novel Artifact Removal Strategy and Spatial Attention-based Multiscale CNN for MI Recognition**

*Authors: Duan Li, Peisen Liu, Yongquan Xia*

**PAGE 283 – 293**

**Paper 32: SFFT-CapsNet: Stacked Fast Fourier Transform for Retina Optical Coherence Tomography Image Classification using Capsule Network**

*Authors: Michael Opoku, Benjamin Asubam Weyori, Adebayo Felix Adekoya, Kwabena Adu*

**PAGE 294 – 306**

**Paper 33: Deep Neural Network-based Detection of Road Traffic Objects from Drone-Captured Imagery Focusing on Road Regions**

*Authors: Hoanh Nguyen*

**PAGE 307 – 314**

**Paper 34: Strengthening Network Security: Evaluation of Intrusion Detection and Prevention Systems Tools in Networking Systems**

*Authors: Wahyu Adi Prabowo, Khusnul Fauziah, Aufa Salsabila Nahrowi, Muhammad Nur Faiz, Arif Wirawan Muhammad*

**PAGE 315 – 324**

**Paper 35: Hybrid Local Search Algorithm for Optimization Route of Travelling Salesman Problem**

*Authors: Muhammad Khahfi Zuhanda, Noriszura Ismail, Rezzy Eko Caraka, Rahmad Syah, Prana Ugiana Gio*

**PAGE 325 – 332**

**Paper 36: A Systematic Literature Review of Computational Studies in Aquaponic System**

*Authors: Khaoula Taji, Ali Sohail, Yassine Taleb Ahmad, Ilyas Ghanimi, Sheeba Ilyas, Fadoua Ghanimi*

**PAGE 333 – 343**

**Paper 37: Cocoa Pods Diseases Detection by MobileNet Confluence and Classification Algorithms**

*Authors: Diarra MAMADOU, Kacoutchy Jean AYIKPA, Abou Bakary BALLO, Brou Médard KOUASSI*

**PAGE 344 – 352**

**Paper 38: Wireless Capsule Endoscopy Video Summarization using Transfer Learning and Random Forests**

*Authors: Parminder Kaur, Rakesh Kumar*

**PAGE 353 – 358**

**Paper 39: Contributed Factors in Predicting Market Values of Loaned Out Players of English Premier League Clubs**

*Authors: Muhammad Daffa Arviano Putra, Deshinta Arrova Dewi, Wahyuningdiah Trisari Putri, Refno Hendrowati, Tri Basuki Kurniawan*

**PAGE 359 – 365**

**Paper 40: Exploring the Challenges and Impacts of Artificial Intelligence Implementation in Project Management: A Systematic Literature Review**

*Authors: Muhammad Irfan Hashfi, Teguh Raharjo*

**PAGE 366 – 376**

**Paper 41: Deep Residual Convolutional Long Short-term Memory Network for Option Price Prediction Problem**

*Authors: Artur Dossatayev, Ainur Manapova, Batyrkhan Omarov*

**PAGE 377 – 387**

**Paper 42: Factors and Models Influencing Value Co-Creation in the Supply Chain of Collection Resources for Library Distribution Providers Under Data Ecology**

*Authors: Xiaoyun Lin*

**PAGE 388 – 397**

**Paper 43: A Flexible Manufacturing System based on Virtual Simulation Technology for Building Flexible Platforms**

*Authors: Zhangchi Sun*

**PAGE 398 – 407**

**Paper 44: Enhanced Plagiarism Detection Through Advanced Natural Language Processing and E-BERT Framework of the Smith-Waterman Algorithm**

*Authors: Franciskus Antonius, Myagmarsuren Orosoo, Aanandha Saravanan K, Indrajit Patra, Prema S*

**PAGE 408 – 416**

**Paper 45: Comparative Study of Machine Learning Algorithms for Phishing Website Detection**

*Authors: Kamal Omari*

**PAGE 417 – 425**

**Paper 46: Study of the Impact of the Internet of Things Integration on Competition Among 3PLs**

*Authors: Kenza Izikki, Mustapha Hlyal, Aziz Ait Bassou, Jamila El Alami*

**PAGE 426 – 434**

**Paper 47: A Framework for Predicting Academic Success using Classification Method through Filter-Based Feature Selection**

*Authors: Dafid, Ermatita, Samsuryadi*

**PAGE 435 – 444**

**Paper 48: A Performance Analysis of Point CNN and Mask R-CNN for Building Extraction from Multispectral LiDAR Data**  
*Authors: Asmaa A. Mandouh, Mahmoud El Nokrashy O. Ali, Mostafa H.A. Mohamed, Lamyaa Gamal EL-Deen Taha, Sayed A. Mohamed*

**PAGE 445 – 452**

**Paper 49: Securing IoT Devices in e-Health using Machine Learning Techniques**

*Authors: Haifa Khaled Alanazi, A. A. Abd El-Aziz, Hedi Hamdi*

**PAGE 453 – 464**

**Paper 50: A Multispectral Ariel Image Stitching using Decorrelation and EEG Signal Extraction Technique**

*Authors: Mukul Manohar S, K N Muralidhara*

**PAGE 465 – 472**

**Paper 51: Texton Tri-alley Separable Feature Merging (TTSFM) Capsule Network for Brain Tumor Detection**

*Authors: Vivian Akoto-Adjepong, Obed Appiah, Peter Appiahene, Patrick Kwabena Mensah*

**PAGE 473 – 483**

**Paper 52: Unraveling Ransomware: Detecting Threats with Advanced Machine Learning Algorithms**

*Authors: Karam Hammadeh, M. Kavitha*

**PAGE 484 – 491**

**Paper 53: K-Means Extensions for Clustering Categorical Data on Concept Lattice**

*Authors: Mohammed Alwersh, László Kovács*

**PAGE 492 – 507**

**Paper 54: Intelligent Heart Disease Prediction System with Applications in Jordanian Hospitals**

*Authors: Mohammad Subhi Al-Batah, Mowafaq Salem Alzboon, Raed Alazaidah*

**PAGE 508 – 517**

**Paper 55: A Novel Approach for Content-based Image Retrieval System using Logical AND and OR Operations**

*Authors: Ranjana Battur, Jagadisha Narayana*

**PAGE 518 – 528**

**Paper 56: Imperative Role of Digital Twin in the Management of Hospitality Services**

*Authors: Ramnarayan, Rajesh Singh, Anita Gehlot, Kapil Joshi, Ashraf Osman Ibrahim, Anas W. Abulfaraj, Faisal Binzagr, Salil Bharany*

**PAGE 529 – 537**

**Paper 57: Design of a Hypermodel using Transfer Learning to Detect DDoS Attacks in the Cloud Security**

*Authors: Marram Amitha, Muktevi Srivenkatesh*

**PAGE 538 – 544**

**Paper 58: Cyberbullying Detection Based on Hybrid Ensemble Method using Deep Learning Technique in Bangla Dataset**

*Authors: Md. Tofael Ahmed, Afroza Sharmin Urmi, Maqsdur Rahman, Abu Zafor Muhammad Touhidul Islam, Dipankar Das, Md. Golam Rashed*

**PAGE 545 – 551**

**Paper 59: SE-RESNET: Monkeypox Detection Model**

*Authors: Krishnan Thiruppathi, Selvakumar K, Vairachilai Shenbagavel*

**PAGE 552 – 558**



**Paper 60: Enhancing Skin Cancer Detection Through an AI-Powered Framework by Integrating African Vulture Optimization with GAN-based Bi-LSTM Architecture**

*Authors: N. V. Rajasekhar Reddy, Araddhana Arvind Deshmukh, Vuda Sreenivasa Rao, Sanjiv Rao Godla, Yousef A. Baker El-Ebiary, Liz Maribel Robladillo Bravo, R. Manikandan*

**PAGE 559 – 572**

**Paper 61: Enhancing Diabetic Retinopathy Detection Through Machine Learning with Restricted Boltzmann Machines**

*Authors: Venkateswara Rao Naramala, B. Anjanee Kumar, Vuda Sreenivasa Rao, Annapurna Mishra, Shaikh Abdul Hannan, Yousef A. Baker El-Ebiary, R. Manikandan*

**PAGE 573 – 585**

**Paper 62: Feline Wolf Net: A Hybrid Lion-Grey Wolf Optimization Deep Learning Model for Ovarian Cancer Detection**

*Authors: Moresh Mukhedkar, Divya Rohatgi, Veera Ankalu Vuyyuru, K V S S Ramakrishna, Yousef A. Baker El-Ebiary, V. Antony Asir Daniel*

**PAGE 586 – 596**

**Paper 63: Utilizing Deep Convolutional Neural Networks and Non-Negative Matrix Factorization for Multi-Modal Image Fusion**

*Authors: Nripendra Narayan Das, Santhakumar Govindasamy, Sanjiv Rao Godla, Yousef A. Baker El-Ebiary, E. Thenmozhi*

**PAGE 597 – 606**

**Paper 64: Hybrid Image Encryption using Non-Adjacent Bits Dynamic Encoding DNA with RSA and Chaotic Systems**

*Authors: Marwa A. Elmenyawy, Nada M. Abdel Aziem*

**PAGE 607 – 620**

**Paper 65: Object Detection and Recognition in Remote Sensing Images by Employing a Hybrid Generative Adversarial Networks and Convolutional Neural Networks**

*Authors: Araddhana Arvind Deshmukh, Mamta Kumari, V.V. Jaya Rama Krishnaiah, Suraj Bandhekar, R. Dharani*

**PAGE 621 – 632**

**Paper 66: AIRA-ML: Auto Insurance Risk Assessment-Machine Learning Model using Resampling Methods**

*Authors: Ahmed Shawky Elbhrawy, Mohamed A. Belal, Mohamed Sameh Hassanein*

**PAGE 633 – 641**

**Paper 67: An MILP-based Lexicographic Approach for Robust Selective Full Truckload Vehicle Routing Problem**

*Authors: Karim EL Bouyahyiouy, Anouar Annouch, Adil Bellabdaoui*

**PAGE 642 – 650**

**Paper 68: Design of Personalized Recommendation and Sharing Management System for Science and Technology Achievements based on WEBSOCKET Technology**

*Authors: Shan Zuo, Kai Xiao, Taitian Mao*

**PAGE 651 – 660**

**Paper 69: Mechatronics Design and Development of T-EVA: Bio-Sensorized Space System for Astronaut's Upper Body Temperature Monitoring During Extravehicular Activities on the Moon and Mars**

*Authors: Paul Palacios, Jose Cornejo, Juan C. Chavez, Carlos Cornejo, Jorge Cornejo, Mariela Vargas, Natalia I. Vargas-Cuentas, Avid Roman-Gonzalez, Julio Valdivia-Silva*

**PAGE 661 – 672**

**Paper 70: Artificial Intelligence-based Volleyball Target Detection and Behavior Recognition Method**

*Authors: Jieli Huang, Wenjun Zou*

**PAGE 673 – 680**

**Paper 71: Deep Learning-based Multiple Bleeding Detection in Wireless Capsule Endoscopy**

*Authors: Ouiem Bchir, Ghaida Ali Alkhudhair, Lena Saleh Alotaibi, Noura Abdulhakeem Almhizea, Sara Mohammed Almuhanha, Shouq Fahad Alzeer*

**PAGE 681 – 687**

**Paper 72: A Novel Feature Fusion for the Classification of Histopathological Carcinoma Images**

*Authors: Salini S Nair, M. Subaji*

**PAGE 688 – 697**

**Paper 73: Deep Conv-LSTM Network for Arrhythmia Detection using ECG Data**

*Authors: Alisher Mukhamefkaly, Zeinel Momyunkulov, Nurgul Kurmanbekkyzy, Batyrkhan Omarov*

**PAGE 698 – 707**

**Paper 74: DetBERT: Enhancing Detection of Policy Violations for Voice Assistant Applications using BERT**

*Authors: Rawan Baalous, Joud Alzahrani, Mariam Ali, Rana Asiri, Eman Nooli*

**PAGE 708 – 715**

**Paper 75: Digital Stethoscope for Early Detection of Heart Disease on Phonocardiography Data**

*Authors: Batyrkhan Omarov, Assyl Tuimebayev, Rustam Abdrakhmanov, Bakytgul Yeskarayeva, Daniyar Sultan, Kanat Aidarov*

**PAGE 716 – 724**

**Paper 76: Predicting the Level of Safety Feeling of Bangladeshi Internet users using Data Mining and Machine Learning**

*Authors: Md. Safiul Alam, Anirban Roy, Partha Protim Majumder, Sharun Akter Khushbu*

**PAGE 725 – 739**

**Paper 77: A Novel Deep Neural Network to Analyze and Monitoring the Physical Training Relation to Sports Activities**

*Authors: Bakhytzhhan Omarov, Nurlan Nurmash, Bauyrzhan Doskarayev, Nagashbek Zhilisbaev, Maxat Dairabayev, Shamurat Orazov, Nurlan Omarov*

**PAGE 740 – 747**

**Paper 78: Hybrid CNN-LSTM Network for Cyberbullying Detection on Social Networks using Textual Contents**

*Authors: Daniyar Sultan, Mateus Mendes, Aray Kassenkhan, Olzhas Akyzbekov*

**PAGE 748 – 756**

**Paper 79: Real-Time Road Surface Damage Detection Framework based on Mask R-CNN Model**

*Authors: Bakhytzhhan Kulambayev, Magzat Nurlybek, Gulnar Astaubayeva, Gulnara Tleuberdiyeva, Serik Zholdasbayev, Abdimukhan Tolep*

**PAGE 757 – 765**

**Paper 80: Osteoporosis Detection and Classification of Femur X-ray Images Through Spectral Domain Analysis using Texture Features**

*Authors: Dhanyavathi A, Veena M B*

**PAGE 766 – 773**

**Paper 81: A Systematic Review on Blockchain Scalability**

*Authors: Asmaa Aldoubaee, Noor Hafizah Hassan, Fiza Abdul Rahim*

**PAGE 774 – 784**

**Paper 82: Machine Learning Techniques for Diabetes Classification: A Comparative Study**

*Authors: Hiri Mustafa, Chrayah Mohamed, Ourdani Nabil, Aknin Noura*

**PAGE 785 – 790**

**Paper 83: An Optimized Survival Prediction Method for Kidney Transplant Recipients**

*Authors: Benita Jose Chalissery, V. Asha*

**PAGE 791 – 797**

**Paper 84: Analyzing RNA-Seq Gene Expression Data for Cancer Classification Through ML Approach**

*Authors: Abdul Wahid, M Tariq Banday*

**PAGE 798 – 810**

**Paper 85: Historical Building 3D Reconstruction for a Virtual Reality-based Documentation**

*Authors: Ahmad Zainul Fanani, Arry Maulana Syarif*

**PAGE 811 – 818**

**Paper 86: Identifying and Prioritizing Digital Transformation Elements Using Fuzzy Analytic Hierarchy Process**

*Authors: Mohammed Hitham M.H, Hatem Elkadi, Neamat El Tazi*

**PAGE 819 – 831**

**Paper 87: Machine Learning based Predictive Modelling of Cybersecurity Threats Utilising Behavioural Data**

*Authors: Ting Tin Tin, Khiew Jie Xin, Ali Aitizaz, Lee Kuok Tiung, Teoh Chong Keat, Hasan Sarwar*

**PAGE 832 – 840**

**Paper 88: Enterprise Marketing Decision: Advertising Click Through Rate Prediction Based on Deep Neural Networks**

*Authors: Luyao Zhan*

**PAGE 841 – 849**

**Paper 89: Design and Development of an Intelligent Rendering System for New Year's Paintings Color Based on B/S Architecture**

*Authors: Zaozao Guo*

**PAGE 850 – 861**

**Paper 90: Using EEG Effective Connectivity Based on Granger Causality and Directed Transfer Function for Emotion Recognition**

*Authors: Weisong Wang, Wenjing Sun*

**PAGE 862 – 868**

**Paper 91: Development of an Image Encryption Algorithm using Latin Square Matrix and Logistics Map**

*Authors: Emmanuel Oluwatobi Asani, Godsfavour Biety-Nwanju, Abidemi Emmanuel Adeniyi, Salil Bharany, Ashraf Osman Ibrahim, Anas W. Abulfaraj, Wamda Nagmeldin*

**PAGE 869 – 877**

**Paper 92: Tampering Detection and Segmentation Model for Multimedia Forensic**

*Authors: Manjunatha S, Malini M Patil, Swetha M D, Prabhu Vijay S S*

**PAGE 878 – 887**

**Paper 93: PRESSNet: Assessment of Building Damage Caused by the Earthquake**

*Authors: Dewa Ayu Defina Audrey Nathania, Alexander Agung Santoso Gunawan, Edy Irwansyah*

**PAGE 888 – 894**

**Paper 94: Group Intelligence Recommendation System based on Knowledge Graph and Fusion Recommendation Model**

*Authors: Chengning Huang, Bo Jing, Lili Jiang, Yuquan Zhu*

**PAGE 895 – 904**

**Paper 95: Statistical Language Model-based Analysis of English Corpora and Literature**

*Authors: Wenwen Chai*

**PAGE 905 – 913**

**Paper 96: Marginal Distribution Algorithm for Feature Model Test Configuration Generation**

*Authors: Mohd Zanes Sahid, Mohd Zainuri Saringat, Mohd Hamdi Irwan Hamzah, Nurezayana Zainal*

**PAGE 914 – 924**

**Paper 97: A QoS-Aware Resource Allocation Method for Internet of Things using Ant Colony Optimization Algorithm and Tabu Search**

*Authors: Shuling YIN, Renping YU*

**PAGE 925 – 934**

**Paper 98: Artificial Rabbits Optimizer with Deep Learning Model for Blockchain-Assisted Secure Smart Healthcare System**

*Authors: Mousa Mohammed Khubrani*

**PAGE 935 – 945**

**Paper 99: A QoS-aware Mechanism for Reducing TCP Retransmission Timeouts using Network Tomography**

*Authors: Jingfu Li*

**PAGE 946 – 954**

**Paper 100: Routing Strategies and Protocols for Efficient Data Transmission in the Internet of Vehicles: A Comprehensive Review**

*Authors: Yijun Xu*

**PAGE 955 – 965**

**Paper 101: Design and Implementation Submarine Cable Object Detection YOLOv4 based with Graphical User Interface (GUI) for Remotely Operated Vehicle (ROV)**

*Authors: Fikri Arif Wicaksana, Eueung Mulyana, Syarif Hidayat, Rahadian Yusuf*

**PAGE 966 – 981**

**Paper 102: Research on Clothing Color Classification Method based on Improved FCM Clustering Algorithm**

*Authors: Jinliang Liu*

**PAGE 982 – 989**

**Paper 103: Next-Generation Intrusion Detection and Prevention System Performance in Distributed Big Data Network Security Architectures**

*Authors: Michael Hart, Rushit Dave, Eric Richardson*

**PAGE 990 – 998**

**Paper 104: Machine Learning for Smart Cities: A Comprehensive Review of Applications and Opportunities**

*Authors: Xiaoning Dou, Weijing Chen, Lei Zhu, Yingmei Bai, Yan Li, Xiaoxiao Wu*

**PAGE 999 – 1016**

**Paper 105: Two Dimensional Deep CNN Model for Vision-based Fingerspelling Recognition System**

*Authors: Zhadra Kozhamkulova, Elmira Nurlybaeva, Leilya Kuntunova, Shirin Amanzholova, Marina Vorogushina, Mukhit Maikotov, Kaden Kenzhekhan*

**PAGE 1017 – 1024**

**Paper 106: Non-contact Respiratory Rate Monitoring Based on the Principal Component Analysis**

*Authors: Hoda El Boussaki, Rachid Latif, Amine Saddik, Zakaria El Khadiri, Hicham El Boujaoui*

**PAGE 1025 – 1030**

**Paper 107: An Improved Genetic Algorithm with Chromosome Replacement and Rescheduling for Task Offloading**

*Authors: Hui Fu, Guangyuan Li, Fang Han, Bo Wang*

**PAGE 1031 – 1039**

**Paper 108: An Efficient Convolutional Neural Network Classification Model for Several Sign Language Alphabets**

*Authors: Ahmed Osman Mahmoud, Ibrahim Ziedan, Amr Ahmed Zamel*

**PAGE 1040 – 1050**

**Paper 109: Enhancing Outdoor Mobility and Environment Perception for Visually Impaired Individuals Through a Customized CNN-based System**

*Authors: Athulya N K, Sivakumar Ramachandran, Neetha George, Ambily N, Linu Shine*

**PAGE 1051 – 1058**

**Paper 110: ArCyb: A Robust Machine-Learning Model for Arabic Cyberbullying Tweets in Saudi Arabia**

*Authors: Khalid T. Mursi, Abdulrahman Y. Almalki, Moayad M. Alshangiti, Faisal S. Alsubaei, Ahmed A. Alghamdi*

**PAGE 1059 – 1067**

**Paper 111: Development of a Touchless Control System for a Clinical Robot with Multimodal User Interface**

*Authors: Julio Alegre Luna, Anthony Vasquez Rivera, Alejandra Loayza Mendoza, Jes´us Talavera S., Andres Montoya A*

**PAGE 1068 – 1075**

**Paper 112: Dual-Level Blind Omnidirectional Image Quality Assessment Network Based on Human Visual Perception**

*Authors: Deyang Liu, Lu Zhang, Lifei Wan, Wei Yao, Jian Ma, Youzhi Zhang*

**PAGE 1076 – 1084**

**Paper 113: A Novel Voice Feature AVA and its Application to the Pathological Voice Detection Through Machine Learning**

*Authors: Abdulrehman Altaf, Hairulnizam Mahdin, Ruhaila Maskat, Shazlyn Milleana Shaharudin, Abdullah Altaf, Awais Mahmood*

**PAGE 1085 – 1092**

**Paper 114: Improved Model for Smoke Detection Based on Concentration Features using YOLOv7tiny**

*Authors: Yuanpan ZHENG, Liwei Niu, Xinxin GAN, Hui WANG, Boyang XU, Zhenyu WANG*

**PAGE 1093 – 1103**

**Paper 115: Virtual Machine Allocation in Cloud Computing Environments using Giant Trevally Optimizer**

*Authors: Hai-yu Zhang*

**PAGE 1104 – 1113**

**Paper 116: Corpus Generation to Develop Amharic Morphological Segmenter**

*Authors: Terefe Feyisa, Seble Hailu*

**PAGE 1114 – 1122**

**Paper 117: A Novel Fingerprint Liveness Detection Method using Empirical Mode Decomposition and Neural Network**

*Authors: Shekun Tong, Chunmeng Lu*

**PAGE 1123 – 1131**

**Paper 118: Providing a Hybrid and Symmetric Encryption Solution to Provide Security in Cloud Data Centers**

*Authors: Desong Shen*

**PAGE 1132 – 1141**

**Paper 119: A Single-Stage Deep Learning-based Approach for Real-Time License Plate Recognition in Smart Parking System**

*Authors: Lina YU, Shaokun LIU*

**PAGE 1142 – 1150**

**Paper 120: Enhancing Decision-Making with Data Science in the Internet of Things Environments**

*Authors: Lei Hu, Yangxia Shu*

**PAGE 1151 – 1162**

**Paper 121: A Fruit Ripeness Detection Method using Adapted Deep Learning-based Approach**

*Authors: Weiwei Zhang*

**PAGE 1163 – 1169**

**Paper 122: A New Method for Intrusion Detection in Computer Networks using Computational Intelligence Algorithms**

*Authors: Yanrong HAO, Shaohui YAN*

**PAGE 1170 – 1184**

**Paper 123: Autism Diagnosis using Linear and Nonlinear Analysis of Resting-State EEG and Self-Organizing Map**

*Authors: Jie Xu, Wenxiao Yang*

**PAGE 1185 – 1193**

**Paper 124: A Composite Noise Removal Network Based on Multi-domain Adaptation**

*Authors: Fan Bai, Pengfei Li, Haoyang Sun, Hui Zhang*

**PAGE 1194 – 1205**

# Data Anomaly Detection in the Internet of Things: A Review of Current Trends and Research Challenges

Min Yang<sup>1\*</sup>, Jiajie Zhang<sup>2</sup>

Department of Information Engineering, Shandong Communication & Media College, Jinan 250200, Shandong, China<sup>1</sup>  
School of Intelligent Transportation, Shandong Technician Institute, Jinan 250200, Shandong, China<sup>2</sup>

**Abstract**—The Internet of Things (IoT) has revolutionized how we interact with the physical world, bringing a new era of connectivity. Billions of interconnected devices seamlessly communicate, generating an unprecedented volume of data. However, the dramatic growth of IoT applications also raises an important issue: the reliability and security of IoT data. Data anomaly detection plays a pivotal role in addressing this critical issue, allowing for identifying abnormal patterns, deviations, and malicious activities within IoT data. This paper discusses the current trends, methodologies, and challenges in data anomaly detection within the IoT domain. In this paper, we discuss the strengths and limitations of various anomaly detection techniques, such as statistical methods, machine learning algorithms, and deep learning methods. IoT data anomaly detection carries unique characteristics and challenges that must be carefully considered. We explore these intricacies, such as data heterogeneity, scalability, real-time processing, and privacy concerns. By delving into these challenges, we provide a holistic understanding of the complexity associated with IoT data anomaly detection, paving the way for more targeted and effective solutions.

**Keywords**—Internet of things; anomaly detection; security; machine learning

## I. INTRODUCTION

The Internet of Things (IoT) is a network of connected objects, systems, and devices that gather, share, and react to data. It facilitates device-to-human communication utilizing sensors, software, and internet connectivity [1]. IoT facilitates diverse applications and services, spanning from intelligent residences and urban environments to industrial automation and healthcare surveillance [2]. By seamlessly integrating physical objects into the digital realm, IoT enhances operational effectiveness, refines decision-making processes, and enables unprecedented levels of automation and connectivity across various facets of our everyday experiences [3]. The IoT structure typically consists of three main layers: perception, network, and application. The perception layer, also known as the sensing layer or physical layer, is the lowest layer of the IoT architecture. It comprises physical devices and sensors that capture data from the physical world. The network layer, also known as the communication layer, facilitates the connection and transmission of data between IoT devices and systems employing Wi-Fi, Bluetooth, Zigbee, cellular networks, or even IoT-specific protocols such as MQTT and CoAP. The application layer is the topmost layer in the IoT-layered structure. This layer utilizes data from IoT devices to provide valuable insights and services. It processes and

analyzes the data for various purposes, including data visualization, decision-making, automation, and control [4, 5].

Smart cities, healthcare, industrial automation, and transportation are among the sectors that have benefited greatly from IoT's rapid growth. Since IoT devices generate huge amounts of data, ensuring the integrity and reliability of this data is crucial [6]. There is a significant threat to security and efficiency in IoT systems due to anomalous data patterns, deviations, and outliers [7]. Detecting data anomalies in the IoT is crucial for several reasons. Firstly, anomalies can indicate system malfunctions, faults, or cyberattacks that may disrupt normal operations or compromise the safety and privacy of individuals and organizations [8]. Early detection of anomalies enables proactive measures and timely responses to mitigate potential risks. Secondly, anomaly detection is crucial in optimizing system performance, enhancing decision-making processes, and ensuring data quality. Organizations can improve operational efficiency, optimize resource allocation, and gain valuable insights from the collected data by identifying unusual patterns or outliers in the data [9, 10].

The significance of data anomaly detection in the IoT lies in its potential to enhance system reliability, security, and overall performance. Through traditional monitoring approaches it identifies critical events, anomalies, or irregularities that may go unnoticed [11]. Machine learning and advanced analytics can identify data anomalies in IoT systems to enable real-time insights and preventive maintenance [12]. This not only boosts operational efficiency within the IoT landscape but also ensures the safety, security, and sustainability of the infrastructure. Given the dynamic nature of IoT deployments and the diverse range of IoT devices and applications, effective data anomaly detection techniques are required. These techniques must be scalable, adaptable, and capable of handling high volumes of data [13].

The detection of abnormal patterns or behaviors within data flows generated by IoT sensors is of utmost importance, especially in fields that largely rely on IoT technology, such as education and agriculture. Within these industries, the seamless integration of devices results in a substantial amount of data that facilitates improvements in operational effectiveness and fosters innovation. Nevertheless, the increase in data volume also exposes these fields to possible vulnerabilities, hence emphasizing the significance of identifying atypical or unsuitable patterns to uphold integrity and ensure security. Through the utilization of sophisticated anomaly detection methods, educators and agricultural practitioners have the ability to protect against malevolent behaviors, deviations, and

anomalies that have the potential to disrupt systems or jeopardize confidential data. This process not only guarantees the dependability of IoT-driven processes but also emphasizes the crucial significance of anomaly detection in strengthening the fundamental aspects of Education and Agriculture. This enables these sectors to effectively utilize the advantages offered by IoT technology while maintaining the integrity of data and achieving operational excellence.

Consequently, there is a growing interest in anomaly detection algorithms, innovative data preprocessing techniques, and integrating anomaly detection with real-time analytics and decision-making systems. In order to deploy IoT systems across various domains reliably and securely, advanced data anomaly detection techniques are needed [14]. This study makes the following major contributions:

- To conduct a comprehensive review of the current trends and techniques used for data anomaly detection in the context of the IoT.
- To identify and analyze the major research challenges and limitations associated with data anomaly detection in IoT environments.
- To explore and evaluate existing anomaly detection methods and algorithms, including statistical approaches, machine learning techniques, and anomaly-scoring mechanisms.
- To investigate the impact of different IoT data characteristics, such as high dimensionality, heterogeneity, and dynamic nature, on the performance of anomaly detection methods.
- To propose potential solutions and strategies for enhancing the accuracy, efficiency, and scalability of data anomaly detection in IoT applications.
- To highlight the open research issues and future directions in the field of data anomaly detection in IoT, providing insights for researchers and practitioners.

The remainder of the paper is organized in the following manner. Data anomaly detection strategies are reviewed in Section II. Challenging problems in IoT data anomaly detection are outlined in Section III. Discussion in Section IV and future research directions are highlighted in Section V. Finally, Section VI concludes the paper.

## II. DATA ANOMALY DETECTION STRATEGIES IN IOT

Anomaly detection in the context of IoT involves identifying unusual or abnormal behavior in the data generated by IoT devices. Statistical methods support IoT anomaly detection by leveraging various statistical techniques to detect deviations from expected patterns [15]. These methods analyze the data collected from IoT devices and apply statistical models to identify anomalies that could indicate potential security breaches, system failures, or other abnormal events [16]. One widely used statistical method for IoT anomaly detection is the use of probability distributions [17]. This approach assumes that the data generated by IoT devices follow a specific probability distribution, such as Gaussian or Poisson distribution. By fitting the observed data to these distributions,

statistical parameters can be estimated, allowing for the identification of anomalies based on deviations from the expected distribution [18]. For example, if the data deviate significantly from the mean or exhibit unusually high or low values, it could indicate the presence of anomalies. Time series analysis is another statistical method commonly employed in IoT anomaly detection [19]. IoT data often exhibit temporal dependencies, where the measurements captured by devices are collected over time. Time series analysis techniques, such as Autoregressive Integrated Moving Average (ARIMA) or exponential smoothing models, can be used to model and forecast the expected behavior of the data. Anomalies are then detected by comparing the observed and predicted values, and any significant deviations from the expected pattern are flagged as anomalies [20].

An anomaly can be described as a data point that exhibits a substantial deviation from the expected behavior within a modeled system. Anomalies are generally regarded as infrequent events or observations that significantly deviate from known patterns of behavior. These aberrations have the potential to occur in a single data point, a particular context or temporal segment, or even over the whole dataset. Anomalies, at their core, are frequently ascribed to extraneous variables, such as sensor faults or external assaults [21]. The main goal of a detection algorithm is to accurately identify occurrences of abnormalities, while also classifying or deducing their root causes. The careful selection of an approximation model that closely matches with the expected behavior of the data is crucial in the domain of binary classification for anomalies. Furthermore, the complexities inherent in various situations frequently require customized detection approaches that are specifically designed for each specific application. Fig. 1 illustrates a visual representation of several abnormalities as examples. The categorization of an IoT anomaly detection approach is derived by integrating the classifications presented in prior scholarly works, including [22]. The categorization of algorithms is determined by their problem-solving technique, application, method type, and algorithmic delay. Fig. 2 provides a visual representation of the four categories, offering a comprehensive and explanatory perspective.

A prevalent categorization of anomalies comprises three primary types: point, contextual, and collective anomalies. Point anomaly pertains to situations where a single data point diverges significantly from the anticipated behavior. An illustrative example involves the detection of credit card fraud [23]. In contextual anomaly, an instance could be regarded as anomalous within a particular context. Comparing multiple perspectives of the same data point might not consistently reveal anomalous behavior. Detection of contextual anomalies hinges upon considering both contextual and behavioral attributes together. For instance, anomalies related to traffic violations differ based on geo-location information [24]. Unlike point or contextual anomalies, the collective anomaly examines the entire dataset. A prime example of this type involves the use of electrocardiograms to monitor and identify anomalies or irregularities in the human heart's functioning [25].



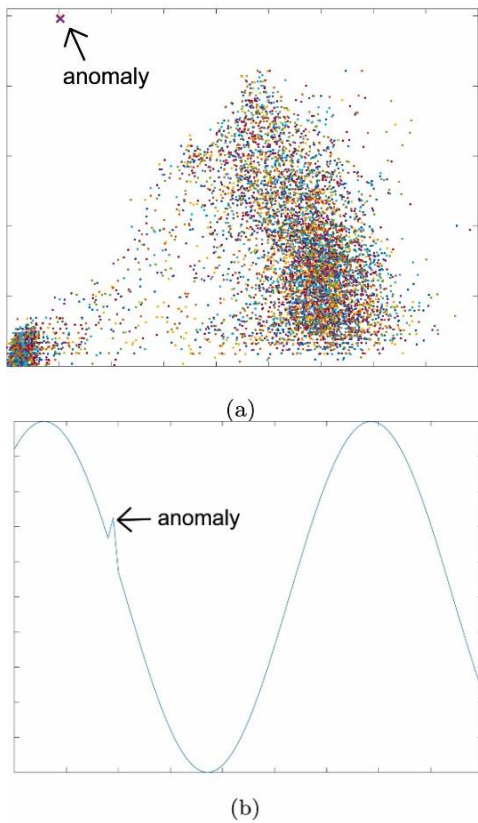


Fig. 1. Visual representations of anomalous occurrences.

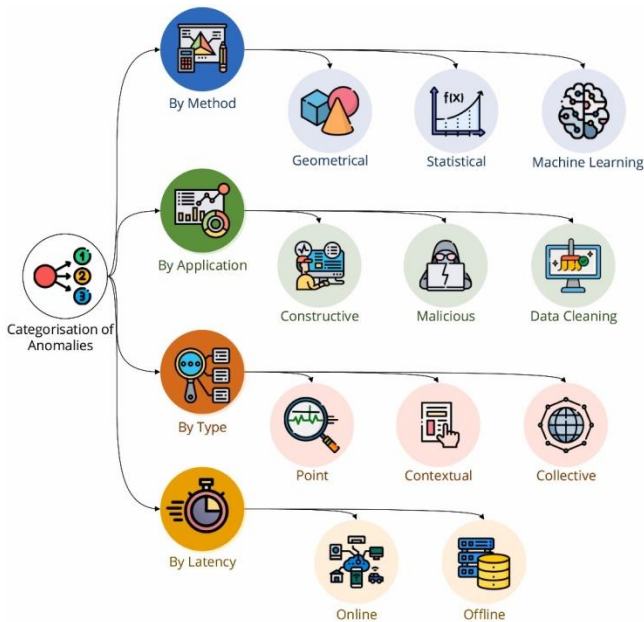


Fig. 2. An overview of anomaly classification.

Anomaly categorization by application can be classified into three distinct routes: constructive, destructive, and data cleaning. Constructive applications are inherently productive and contribute value to various domains. Examples include monitoring the daily activities of the elderly to prevent falls using image descriptors [26]. These applications encompass evaluating the performance of classifiers like multilayer

perceptron (MLP), k-nearest neighbors (k-NN), and support vector machines (SVM). Another instance involves the utilization of reinforcement learning by Lu, et al. [27] for diverse unmanned aerial vehicle (UAV) applications, such as smart farming. Additionally, Nguyen, et al. [28] employ a federated learning approach for smart home applications. Destructive applications are devised to disrupt regular operations, often for dubious financial gains or with intentions to inflict harm upon networks, application data flowing through IoT networks, or critical business practices. These applications have a detrimental impact on society. For instance, Alsheikh, et al. [29] conducted a survey on IoT cyberattacks, shedding light on the latest advancements in IoT security. Solutions to counter such applications, such as RAPPER [30] and NBaIoT [31] employing autoencoders (AEs), focus on prevention or preemptive measures taken before an illicit incident, as well as detection or actions executed after an incident. Data cleaning or data cleansing applications, like DeepAnT [32], employ deep convolutional neural networks (CNNs) to eliminate unwanted data spikes and sensor noise from input signals. These applications play a pivotal role in enhancing the quality of data used in various contexts.

The latency and scalability characteristics of a detection algorithm play a pivotal role in determining its execution timeline, whether it operates on the fly during data collection or at a later storage stage. Online algorithms operate in a serial manner, processing information either one data point at a time or within a window. These algorithms function without requiring access to the entire input dataset. Traditional online methods encompass geometrical and statistical approaches, including distance-based, density-based, and deviation-based techniques. Illustrative examples of online methods are the IoT-Keeper by Hafeez, et al. [33], employing fuzzy C-means, and Bosman, et al. [34] adopt an ensemble approach. Offline algorithms, in contrast, have access to the complete dataset. These algorithms tend to be computationally intensive and sophisticated, aimed at solving complex problems within a reasonable timeframe. It is important to highlight that recent advancements have blurred the distinction between online and offline methods. For instance, [35] utilizes LSTM and Gaussian Naive Bayes, along with other aforementioned models, to perform model training offline and subsequently deploy the model online. This integration allows for more flexibility in the deployment process.

The methods employed can be categorized into geometrical, statistical, or machine learning approaches. Geometrical methods operate under the premise that when employing distance-based or density-based strategies to depict a dataset, the anticipated and anomalous data points become distinguishable. Within a dataset, the underlying principle of isolation or density-based techniques revolves around the notion that anomalies tend to manifest within sparse regions. These techniques utilize a threshold, denoted as 't', either statically or dynamically on the calculated distance 'd' to classify anomalies. This threshold-driven classification is represented by the following equation:

$$d = \begin{cases} < t, & \text{Normal (under threshold)} \\ > t, & \text{Anomaly (above threshold)} \end{cases}$$

Statistical methods, exemplified by the minimal volume technique in [36], aim to model normal data patterns through mathematical models and distributions. The minimal volume approach constructs an n-dimensional simplex around the provided data cloud (considered as ground truth). The objective is to minimize the volume enclosed by the simplex while maximizing the inclusion of ground truth data points. Any data point that does not conform to the simplex is classified as an anomaly. Another example is the forecasting technique known as exponential smoothing [37], which predicts future data points using previous data and a smoothing parameter. Anomalous data detected via statistical methods are those that significantly diverge from the established model.

The third subcategory encompasses machine learning and deep learning models, which have seen an uptick in publication frequency in recent years. The choice of model is contingent on the inherent characteristics of the supplied data [38]. For instance, when dealing with sequential data inputs like audio, video, and time series, models like long short-term memory (LSTM) and transformer models tend to be preferred [39]. Conversely, non-sequential data types, such as image inputs, align well with convolutional neural networks (CNN) and autoencoders (AE) [40, 41]. These algorithms endeavor to discern between normal and anomalous behaviors by establishing a decision boundary. Examples include the utilization of SVM classifiers [42] to delineate such boundaries or employing LSTM networks [35] for future value forecasting in streaming data [43]. The nature of the task dictates whether these approaches fall under the categories of supervised, semi-supervised, self-supervised, or completely unsupervised learning [22, 44], depending on the availability of training labels.

Machine learning algorithms also are crucial to statistical methods for IoT anomaly detection. These algorithms learn patterns and relationships from the data and use them to classify normal and abnormal events. Supervised learning techniques, such as SVM or random forests, excel in scenarios where labeled data is available. By training on labeled data, where anomalies are specifically identified, these techniques generate models capable of automatically detecting anomalies in new, unlabeled data. On the other hand, unsupervised learning techniques, including clustering or outlier detection algorithms, prove valuable in identifying abnormal data points without the need for labeled training data.

Furthermore, statistical methods for IoT anomaly detection often involve thresholds or rule-based approaches. These methods establish predefined thresholds or rules based on the statistical properties of the data. Any data point that exceeds these thresholds or violates the predefined rules is considered an anomaly. For example, if the temperature readings from a temperature sensor exceed a certain predefined range, it could indicate a malfunction or abnormal condition. Statistical methods for IoT anomaly detection encompass a range of techniques, including probability distributions, time series analysis, machine learning algorithms, and threshold-based approaches. By leveraging statistical models and algorithms, these methods can effectively detect anomalies in the data generated by IoT devices, enabling proactive monitoring, early

detection of abnormal events, and mitigation of potential risks in various IoT applications [45].

#### A. Machine Learning Algorithms for Anomaly Detection

Machine learning algorithms enable the automated and efficient detection of abnormal events in the vast amount of data generated by IoT devices [46, 47]. As IoT systems become increasingly complex and interconnected, traditional rule-based or threshold-based approaches may not be sufficient to capture anomalies' diverse and evolving patterns. Machine learning algorithms can learn from historical data, identify hidden patterns, and adapt to changing conditions, making them well-suited for IoT anomaly detection. One key advantage of machine learning algorithms in IoT anomaly detection is their ability to handle large-scale and heterogeneous data [48]. IoT environments generate various data types, including sensor readings, network traffic data, and system logs. Machine learning algorithms can process and analyze this data to identify abnormal patterns that may indicate security breaches, system failures, or other abnormal behavior. These algorithms can handle the high volume, velocity, and variety of IoT data, making them scalable and applicable to real-time monitoring and analysis [49].

Machine learning algorithms are also capable of detecting anomalies that may not be recognized explicitly or anticipated in advance [50]. The capabilities of machine learning algorithms differ from those of rule-based approaches because they can learn from historical data and detect anomalies that may not be apparent to humans. This allows for proactive anomaly detection and early warning of potential issues, reducing the risk of system downtime or security breaches. Machine learning algorithms also offer the advantage of adaptability to changing IoT environments. As IoT systems evolve and new anomalies emerge, machine learning algorithms can continuously learn and update their models to capture these changes. This adaptability is crucial in dynamic IoT environments where anomalies manifest in various forms and evolve over time. By continuously analyzing and updating their models, machine learning algorithms can effectively detect and respond to emerging anomalies, ensuring the reliability and security of IoT systems.

Automated anomaly detection in the realm of IoT is made possible through the utilization of machine learning algorithms. This approach significantly reduces the need for manual inspections and analysis [51]. Manual analysis of IoT data is known to be a time-consuming, error-prone, and inefficient process, particularly in large-scale deployments. By employing machine learning algorithms, data streams from IoT devices can be continuously and efficiently monitored in real time. This empowers human operators to focus their attention on more critical tasks, such as investigating anomalies, taking appropriate actions, or fine-tuning the anomaly detection system. In IoT anomaly detection, machine learning algorithms are indispensable for handling large-scale, heterogeneous data, detecting previously unseen anomalies, adjusting to changing environments, and automating the process. By leveraging machine learning, IoT systems become more secure, reliable, and efficient by proactively detecting and mitigating abnormal events.

### B. Deep Learning Algorithms for Anomaly Detection

The power of neural networks allows deep learning algorithms to detect IoT anomalies by analyzing the data generated by IoT devices and learning complex patterns and representations [52]. Deep learning algorithms, specifically deep neural networks (DNNs), have shown impressive performance in a variety of domains, such as computer vision, natural language processing, and speech recognition. Their ability to automatically extract hierarchical features and model intricate relationships makes them well-suited for detecting anomalies in IoT data. One key advantage of deep learning algorithms in IoT anomaly detection is their ability to handle high-dimensional and unstructured data. IoT environments generate vast amounts of data, often in the form of images, sensor readings, or textual information. Deep learning algorithms can effectively process and analyze this data, capturing subtle and nuanced patterns that may indicate anomalies. Convolutional neural networks (CNNs) excel at analyzing images or sensor data, while recurrent neural networks (RNNs) can handle sequential or time series data. These architectures enable deep learning algorithms to learn highly relevant representations for anomaly detection.

Another crucial aspect of deep learning algorithms is their ability to automatically learn from data without relying on explicit feature engineering. Traditional machine learning algorithms often require manual extraction of relevant features, which can be time-consuming and challenging, especially in the context of IoT data. Deep learning algorithms can autonomously learn and extract relevant features directly from raw data, alleviating the need for extensive domain knowledge and manual feature engineering. This enables them to uncover intricate and non-linear relationships in the data, improving the accuracy and robustness of anomaly detection. Furthermore, deep learning algorithms offer the advantage of transfer learning and knowledge sharing across different IoT domains. Pretrained deep neural networks, trained on large-scale datasets from other domains, can be fine-tuned and adapted to specific IoT anomaly detection tasks. This knowledge transfer allows deep learning algorithms to leverage the learned representations and patterns from other domains, even when

labeled training data is limited or unavailable in the IoT domain. Transfer learning facilitates faster model convergence, improves generalization, and enhances anomaly detection performance in IoT environments.

Deep learning algorithms also exhibit the potential for anomaly detection in real-time or streaming IoT data. Recurrent neural networks, such as LSTM or gated recurrent units (GRU), are well-suited for modeling sequential dependencies in time series data. This makes them effective for detecting anomalies in streaming IoT data, where anomalies can occur in real-time. By analyzing the temporal patterns and dependencies in the data, deep learning algorithms can provide timely detection and response to abnormal events, enabling proactive monitoring and mitigation. Deep learning algorithms, with their ability to handle high-dimensional data, automatically learn relevant features, facilitate transfer learning, and analyze sequential dependencies, are instrumental in IoT anomaly detection. By leveraging deep neural networks, IoT systems can effectively detect anomalies in complex and diverse data generated by IoT devices. The role of deep learning algorithms extends to enhancing the security, reliability, and operational efficiency of IoT systems by enabling proactive anomaly detection and timely mitigation of abnormal events.

### C. Comparative Analysis of the Different Techniques

Table I presents a side-by-side comparison of the machine and deep learning algorithms for IoT data anomaly detection. SVM is known for its high accuracy and effectiveness in handling linearly separable data. It is robust against overfitting and can handle high-dimensional data. However, SVMs can be computationally intensive for large datasets, and selecting appropriate kernel functions requires careful consideration. Random Forests offer high accuracy and are robust against overfitting. They handle high-dimensional data well and provide feature importance rankings. However, they are less interpretable compared to individual decision trees. k-NN is a simple and intuitive algorithm that detects local anomalies. It is non-parametric and adaptive, making it suitable for handling noisy data. However, k-NN is sensitive to the choice of distance metric and requires careful selection of the value for k.

TABLE I. AN OVERVIEW OF THE MACHINE AND DEEP LEARNING ALGORITHMS FOR IOT DATA ANOMALY DETECTION

Algorithm	Performance	Strengths	Limitations	References
Support Vector Machines (SVM)	High accuracy Effective for linearly separable data	Robust against overfitting Can handle high-dimensional data	Computationally intensive for large datasets Requires careful selection of kernel functions	[42, 53-59]
Random Forests (RF)	High accuracy Robust against overfitting	Handles high-dimensional data Provides feature importance rankings	Less interpretable compared to individual decision trees	[8, 60-62]
k-Nearest Neighbors (k-NN)	Simple and intuitive Effective for local anomalies	Non-parametric and adaptive Handles noisy data	Sensitive to the choice of distance metric Requires careful selection of k value	[7, 63-66]
Recurrent Neural Networks (RNN)	Captures sequential dependencies in time series data	Handles variable-length sequences Suitable for streaming data	Can suffer from vanishing/exploding gradients Computationally intensive training	[67-74]
Long Short-Term Memory (LSTM)	Captures long-term dependencies in sequential data	Robust against vanishing gradients Suitable for modeling time series data	Requires more training time compared to traditional RNNs	[75-77]
Convolutional Neural Networks (CNN)	Effective for image or sensor data analysis	Automatically learns hierarchical features Robust to spatial variations	It may require large amounts of labeled training data Computationally intensive for large images	[78-81]

Recurrent Neural Networks (RNNs) capture sequential dependencies in time series data, making them suitable for IoT anomaly detection. They can handle variable-length sequences and are well-suited for streaming data. However, RNNs can suffer from vanishing or exploding gradients during training and can be computationally intensive. LSTM networks are a type of RNN that can capture long-term dependencies in sequential data. They are robust against vanishing gradients and are suitable for modeling time series data. However, LSTM networks generally require more training time compared to traditional RNNs. Convolutional Neural Networks (CNNs) are particularly effective for analyzing image or sensor data in IoT applications. They automatically learn hierarchical features from the data and are robust to spatial variations. CNNs can capture local patterns and spatial dependencies, making them suitable for anomaly detection in image-based IoT data. However, CNNs often require large amounts of labeled training data to achieve optimal performance. Training large CNN models can also be computationally intensive, especially when dealing with high-resolution images or large-scale datasets.

### III. CHALLENGES IN DATA ANOMALY DETECTION FOR IOT

Data anomaly detection in IoT is challenging due to the unique characteristics of IoT data and the constraints imposed by IoT environments. Some of the key challenges are as follows:

- **High dimensionality:** IoT data is often high-dimensional, consisting of multiple sensors, devices, and data sources. This high dimensionality increases the complexity of anomaly detection, as the algorithms need to handle a large number of features and capture complex relationships between them. Dimensionality reduction techniques may be required to mitigate this challenge.
- **Scalability:** IoT systems generate massive amounts of data in real-time. Anomaly detection algorithms must scale to handle the high data volume and velocity. Processing such large-scale data in real-time requires efficient algorithms and infrastructure capable of handling the computational and storage demands.
- **Imbalanced data:** IoT datasets often suffer from imbalanced class distributions, where the number of normal instances significantly outweighs the number of anomalies. This imbalance can lead to biased models that favor the majority class and fail to detect anomalies accurately. Specialized techniques, such as oversampling or undersampling, must address this challenge and improve the detection of rare anomalies.
- **Concept drift:** IoT environments are dynamic and subject to concept drift, where the statistical properties of the data change over time. Anomaly detection models trained on historical data may become less effective when faced with new data patterns. Continuous model updating and adaptation are necessary to cope with concept drift and ensure the detection of evolving anomalies.

- **Lack of labeled data:** Anomaly detection typically requires labeled data for training supervised learning algorithms. However, acquiring labeled data for anomalies can be challenging in IoT settings, as anomalies are rare and may not be explicitly labeled. Obtaining a sufficient amount of accurately labeled data for training can be a significant obstacle, necessitating the exploration of unsupervised or semi-supervised techniques.
- **Privacy and security:** IoT data often contains sensitive information, making privacy and security crucial concerns. Anomaly detection algorithms must operate in a privacy-preserving manner, ensuring that sensitive data is not exposed or compromised during the detection process. This requires carefully designing algorithms and techniques to balance anomaly detection accuracy with privacy protection.
- **Real-time detection:** Many IoT applications require real-time anomaly detection for timely response and mitigation. Achieving real-time detection poses challenges due to the computational complexity of certain algorithms and the need to process and analyze data in near real-time. Efficient algorithms and scalable infrastructure are necessary to enable real-time anomaly detection in IoT environments.
- **Interpretability:** Understanding why a certain instance is flagged as an anomaly is important for effective anomaly management and decision-making. However, some advanced machine learning and deep learning algorithms, while powerful in detecting anomalies, may lack interpretability. Balancing accuracy and interpretability becomes crucial, especially in applications requiring explainability.

Addressing the mentioned challenges in data anomaly detection for IoT requires the development of innovative algorithms, techniques, and frameworks for detecting high-dimensional, streaming data effectively, adjusting to dynamic environments, maintaining privacy, and enabling detection in real-time. Such solutions should also be energy efficient and easily scalable to accommodate large-scale networks. Finally, they should be able to detect anomalies caused by malicious activities, natural phenomena, and human errors.

### IV. DISCUSSION

Various case studies demonstrate the diverse applications of IoT data anomaly detection across industries, ranging from manufacturing and home security to healthcare. Organizations can achieve improved operational efficiency, enhanced security, and proactive decision-making in various IoT-enabled environments by leveraging anomaly detection algorithms. Table II shows an overview of case studies of IoT data anomaly detection. IoT sensors are deployed across machinery and equipment in a manufacturing plant to collect data on parameters such as temperature, vibration, and energy consumption. Anomaly detection algorithms are applied to this data to identify deviations from normal behavior that may indicate potential failures or malfunctions. By detecting anomalies in real-time, maintenance teams can proactively

schedule repairs or replacement of components before costly breakdowns occur. This approach is employed by companies like General Electric (GE) for their industrial IoT applications, resulting in shorter downtime, longer equipment lifetime, and reduced costs.

TABLE II. CASE STUDIES OF IoT DATA ANOMALY DETECTION

Industry	Case study	Key benefits	References
Manufacturing	Predictive maintenance	Reduced downtime Increased equipment lifespan Cost savings	[82-85]
Smart home	Security	Improved home security Real-time anomaly detection Mitigation of potential security risks	[86-89]
Healthcare	Patient monitoring	Early detection of health issues Personalized healthcare monitoring Timely intervention	[90-93]

Smart homes use IoT devices, such as cameras, motion sensors, doors, and windows, to generate data on activities and events within the home. Anomaly detection algorithms are applied to this data to identify abnormal behaviors or potential intrusions. This allows homeowners to receive real-time alerts and take appropriate actions to mitigate security risks. Companies like Ring and Nest have implemented IoT data anomaly detection techniques in their smart home security systems, providing homeowners with improved security and peace of mind. IoT devices and wearables in healthcare generate vast amounts of patient data, including vital signs, activity levels, and medication adherence. Anomaly detection algorithms are applied to this data to identify deviations from normal patterns, indicating potential health issues or abnormal behavior. Healthcare providers can receive alerts and take timely interventions, leading to early detection of health issues and personalized patient care. Companies like Philips and Medtronic utilize IoT data anomaly detection in their healthcare monitoring solutions to improve patient outcomes and enhance healthcare delivery.

## V. FUTURE RESEARCH DIRECTIONS

Future research in IoT data anomaly detection is expected to address several key challenges and explore novel techniques to improve the effectiveness and efficiency of anomaly detection in IoT systems. Here are some potential research directions:

- **Real-time and edge-based anomaly detection:** As the IoT ecosystem continues to grow, there is a need for more real-time and edge-based anomaly detection methods. Research efforts will aim to develop lightweight algorithms and models to efficiently process and analyze IoT data at the edge, reducing latency and enabling timely anomaly detection and response.
- **Robustness to evolving IoT environments:** IoT environments are dynamic, with device changes, data distributions, and system configurations. Future

research will focus on developing anomaly detection techniques that adapt to evolving IoT environments. This includes techniques for transfer learning, online learning, and incremental learning, allowing anomaly detection models to learn and adapt to new patterns and anomalies continuously.

- **Multi-modal anomaly detection:** IoT systems generate data from diverse sources, including sensors, images, audio, and video streams. Future research will explore multi-modal anomaly detection techniques that can effectively integrate and analyze data from different modalities to detect complex anomalies that may not be apparent when analyzing each modality individually.
- **Explainable AI for anomaly detection:** Explainability and interpretability are critical for building trust and understanding in anomaly detection systems. Future research will focus on developing explainable AI techniques for anomaly detection in IoT data. This includes methods to provide interpretable explanations for detected anomalies, visualizations of anomaly patterns, and feature importance analysis to enhance the transparency and usability of anomaly detection models.
- **Privacy-preserving anomaly detection:** IoT data often contain sensitive and personal information. Future research will explore privacy-preserving anomaly detection techniques that can detect anomalies without compromising the privacy of individuals or revealing sensitive data. This includes techniques such as federated learning, secure multi-party computation, and differential privacy to ensure data privacy and security in anomaly detection processes.
- **Adversarial anomaly detection:** As IoT systems become more interconnected and susceptible to attacks; future research will investigate adversarial anomaly detection techniques. These techniques aim to detect anomalies caused by malicious activities, such as data poisoning or evasion attacks. Research efforts will focus on developing robust anomaly detection models to detect and mitigate adversarial attacks on IoT data.

By addressing these research directions, IoT data anomaly detection can advance to effectively handle the complexities and challenges of large-scale IoT systems, leading to more reliable anomaly detection, enhanced security, and improved operational efficiency.

## VI. CONCLUSION

Data anomaly detection plays a crucial role in the IoT ecosystem, enabling the detection of abnormal behavior, potential failures, and security breaches in IoT systems. This paper comprehensively reviews current trends and research challenges in IoT data anomaly detection. We have discussed utilizing machine learning and deep learning algorithms, such as ensemble methods, RNNs, and CNNs, in IoT anomaly detection. These algorithms offer advanced capabilities in

handling IoT data's complexity and high dimensionality, leading to more accurate and efficient anomaly detection. Additionally, unsupervised learning approaches and real-time processing have emerged as prominent trends, enabling the detection of anomalies without the need for labeled data and facilitating timely responses to detected anomalies.

Furthermore, integrating multiple data sources and pursuing explainable AI techniques have been identified as important trends in IoT data anomaly detection. By leveraging diverse sources of IoT data and providing interpretable explanations for detected anomalies, organizations can enhance anomaly detection systems' reliability, usability, and trustworthiness. However, several research challenges remain in the field. These include the development of real-time and edge-based anomaly detection methods, addressing the robustness of anomaly detection models in evolving IoT environments, and exploring multi-modal anomaly detection techniques. Privacy-preserving and adversarial anomaly detection are crucial areas requiring further research to ensure data privacy, security, and resilience against malicious activities.

#### REFERENCES

- [1] B. Pourghebleh, N. Hekmati, Z. Davoudnia, and M. Sadeghi, "A roadmap towards energy - efficient data fusion methods in the Internet of Things," *Concurrency and Computation: Practice and Experience*, p. e6959, 2022.
- [2] A. Zhu, M. Ma, S. Guo, and Y. Yang, "Adaptive Access Selection Algorithm for Multi-Service in 5G Heterogeneous Internet of Things," *IEEE Transactions on Network Science and Engineering*, vol. 9, no. 3, pp. 1630-1644, 2022.
- [3] B. Pourghebleh, K. Wakil, and N. J. Navimipour, "A comprehensive study on the trust management techniques in the Internet of Things," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 9326-9337, 2019.
- [4] 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.
- [5] M. Younan, E. H. Houssein, M. Elhoseny, and A. A. Ali, "Challenges and recommended technologies for the industrial internet of things: A comprehensive review," *Measurement*, vol. 151, p. 107198, 2020.
- [6] B. Pourghebleh and V. Hayyolalam, "A comprehensive and systematic review of the load balancing mechanisms in the Internet of Things," *Cluster Computing*, pp. 1-21, 2019.
- [7] A. M. F. Al-Sammarrie and M. Çevik, "Anomaly detection of web traffic between IoT Devices," in *2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*, 2022: IEEE, pp. 1-3.
- [8] A. Alzahrani, T. Baabdullah, and D. B. Rawat, "Attacks and Anomaly Detection in IoT Network Using Machine Learning," in *HCI International 2021-Posters: 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings, Part II 23*, 2021: Springer, pp. 465-472.
- [9] H. Liu, C. Zhong, A. Alnusair, and S. R. Islam, "FAIXID: A framework for enhancing ai explainability of intrusion detection results using data cleaning techniques," *Journal of network and systems management*, vol. 29, no. 4, p. 40, 2021.
- [10] H. Hoorfar, N. Taheri, H. Kosarirad, and A. Bagheri, "Efficiently Guiding K-Robots Along Pathways with Minimal Turns," *EAI Endorsed Transactions on AI and Robotics*, vol. 2, 2023.
- [11] I. Martins, J. S. Resende, P. R. Sousa, S. Silva, L. Antunes, and J. Gama, "Host-based IDS: A review and open issues of an anomaly detection system in IoT," *Future Generation Computer Systems*, 2022.
- [12] N. Moustafa, N. Koroniotis, M. Keshk, A. Y. Zomaya, and Z. Tari, "Explainable Intrusion Detection for Cyber Defences in the Internet of Things: Opportunities and Solutions," *IEEE Communications Surveys & Tutorials*, 2023.
- [13] L. Yang and A. Shami, "IoT data analytics in dynamic environments: From an automated machine learning perspective," *Engineering Applications of Artificial Intelligence*, vol. 116, p. 105366, 2022.
- [14] E. Şengönül, R. Samet, Q. Abu Al-Haija, A. Alqahtani, B. Alturki, and A. A. Alsulami, "An Analysis of Artificial Intelligence Techniques in Surveillance Video Anomaly Detection: A Comprehensive Survey," *Applied Sciences*, vol. 13, no. 8, p. 4956, 2023.
- [15] Y. Liu, H. Wang, X. Zheng, and L. Tian, "An efficient framework for unsupervised anomaly detection over edge-assisted internet of things," *ACM Transactions on Sensor Networks*, 2023.
- [16] W. Jia, R. M. Shukla, and S. Sengupta, "Anomaly detection using supervised learning and multiple statistical methods," in *2019 18th IEEE International Conference On Machine Learning and Applications (ICMLA)*, 2019: IEEE, pp. 1291-1297.
- [17] S. Maleki, S. Maleki, and N. R. Jennings, "Unsupervised anomaly detection with LSTM autoencoders using statistical data-filtering," *Applied Soft Computing*, vol. 108, p. 107443, 2021.
- [18] J. Pei, K. Zhong, M. A. Jan, and J. Li, "Personalized federated learning framework for network traffic anomaly detection," *Computer Networks*, vol. 209, p. 108906, 2022.
- [19] A. A. Cook, G. Mısırlı, and Z. Fan, "Anomaly detection for IoT time-series data: A survey," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 6481-6494, 2019.
- [20] J. E. Zhang, D. Wu, and B. Boulet, "Time series anomaly detection for smart grids: A survey," in *2021 IEEE Electrical Power and Energy Conference (EPEC)*, 2021: IEEE, pp. 125-130.
- [21] H. Hoorfar, H. Kosarirad, N. Taheri, F. Fathi, and A. Bagheri, "Concealing Robots in Environments: Enhancing Navigation and Privacy through Stealth Integration," *EAI Endorsed Transactions on AI and Robotics*, vol. 2, 2023.
- [22] M. Fahim and A. Sillitti, "Anomaly detection, analysis and prediction techniques in iot environment: A systematic literature review," *IEEE Access*, vol. 7, pp. 81664-81681, 2019.
- [23] P. Srikanth, "An efficient approach for clustering and classification for fraud detection using bankruptcy data in IoT environment," *International Journal of Information Technology*, vol. 13, no. 6, pp. 2497-2503, 2021.
- [24] S. Asoba, S. Supekar, T. Tonde, and J. A. Siddiqui, "Advanced traffic violation control and penalty system using IoT and image processing techniques," in *2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, 2020: IEEE, pp. 554-558.
- [25] H. Li and P. Boulanger, "A survey of heart anomaly detection using ambulatory electrocardiogram (ECG)," *Sensors*, vol. 20, no. 5, p. 1461, 2020.
- [26] Y. M. Galvão, V. A. Albuquerque, B. J. Fernandes, and M. J. Valença, "Anomaly detection in smart houses: Monitoring elderly daily behavior for fall detecting," in *2017 IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, 2017: IEEE, pp. 1-6.
- [27] H. Lu, Y. Li, S. Mu, D. Wang, H. Kim, and S. Serikawa, "Motor anomaly detection for unmanned aerial vehicles using reinforcement learning," *IEEE internet of things journal*, vol. 5, no. 4, pp. 2315-2322, 2017.
- [28] T. D. Nguyen, S. Marchal, M. Miettinen, H. Fereidooni, N. Asokan, and A.-R. Sadeghi, "DİoT: A federated self-learning anomaly detection system for IoT," in *2019 IEEE 39th International conference on distributed computing systems (ICDCS)*, 2019: IEEE, pp. 756-767.
- [29] M. Alsheikh, L. Konieczny, M. Prater, G. Smith, and S. Uludag, "The state of IoT security: Unequivocal appeal to cybercriminals, onerous to defenders," *IEEE Consumer Electronics Magazine*, vol. 11, no. 3, pp. 59-68, 2021.
- [30] M. Alam, S. Sinha, S. Bhattacharya, S. Dutta, D. Mukhopadhyay, and A. Chattopadhyay, "Rapper: Ransomware prevention via performance counters," *arXiv preprint arXiv:2004.01712*, 2020.
- [31] Y. Meidan et al., "N-baiot—network-based detection of iot botnet attacks using deep autoencoders," *IEEE Pervasive Computing*, vol. 17, no. 3, pp. 12-22, 2018.

- [32] M. Munir, S. A. Siddiqui, A. Dengel, and S. Ahmed, "DeepAnT: A deep learning approach for unsupervised anomaly detection in time series," *Ieee Access*, vol. 7, pp. 1991-2005, 2018.
- [33] I. Hafeez, M. Antikainen, A. Y. Ding, and S. Tarkoma, "IoT-KEEPER: Detecting malicious IoT network activity using online traffic analysis at the edge," *IEEE Transactions on Network and Service Management*, vol. 17, no. 1, pp. 45-59, 2020.
- [34] H. H. Bosman, G. Iacca, A. Tejada, H. J. Wörtche, and A. Liotta, "Ensembles of incremental learners to detect anomalies in ad hoc sensor networks," *ad hoc networks*, vol. 35, pp. 14-36, 2015.
- [35] D. Wu, Z. Jiang, X. Xie, X. Wei, W. Yu, and R. Li, "LSTM learning with Bayesian and Gaussian processing for anomaly detection in industrial IoT," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 8, pp. 5244-5253, 2019.
- [36] C. O'Reilly, A. Gluhak, and M. A. Imran, "Distributed anomaly detection using minimum volume elliptical principal component analysis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 9, pp. 2320-2333, 2016.
- [37] S. Mahajan, L.-J. Chen, and T.-C. Tsai, "Short-term PM2. 5 forecasting using exponential smoothing method: A comparative analysis," *Sensors*, vol. 18, no. 10, p. 3223, 2018.
- [38] A. S. Charles, "Interpreting deep learning: The machine learning roschach test?," *arXiv preprint arXiv:1806.00148*, 2018.
- [39] Z. Chen, D. Chen, X. Zhang, Z. Yuan, and X. Cheng, "Learning graph structures with transformer for multivariate time-series anomaly detection in IoT," *IEEE Internet of Things Journal*, vol. 9, no. 12, pp. 9179-9189, 2021.
- [40] A. Ukil, S. Bandyopadhyay, C. Puri, and A. Pal, "IoT healthcare analytics: The importance of anomaly detection," in *2016 IEEE 30th international conference on advanced information networking and applications (AINA)*, 2016: IEEE, pp. 994-997.
- [41] C. Han and X. Fu, "Challenge and Opportunity: Deep Learning-Based Stock Price Prediction by Using Bi-Directional LSTM Model," *Frontiers in Business, Economics and Management*, vol. 8, no. 2, pp. 51-54, 2023.
- [42] K. Yang, S. Kpotufe, and N. Feamster, "An efficient one-class SVM for anomaly detection in the Internet of Things," *arXiv preprint arXiv:2104.11146*, 2021.
- [43] M. Dunne, G. Gracioli, and S. Fischmeister, "A comparison of data streaming frameworks for anomaly detection in embedded systems," in *Proceedings of the 1st International Workshop on Security and Privacy for the Internet-of-Things (IoTSec)*, Orlando, FL, USA, 2018.
- [44] 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.
- [45] R. Al-amri, R. K. Murugesan, M. Man, A. F. Abdulateef, M. A. Al-Sharaf, and A. A. Alkahtani, "A review of machine learning and deep learning techniques for anomaly detection in IoT data," *Applied Sciences*, vol. 11, no. 12, p. 5320, 2021.
- [46] G. Han, J. Tu, L. Liu, M. Martinez-Garcia, and Y. Peng, "Anomaly detection based on multidimensional data processing for protecting vital devices in 6G-enabled massive IIoT," *IEEE Internet of Things Journal*, vol. 8, no. 7, pp. 5219-5229, 2021.
- [47] 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.
- [48] X. Sáez-de-Cámara, J. L. Flores, C. Arellano, A. Urbieto, and U. Zurutuza, "Clustered Federated Learning Architecture for Network Anomaly Detection in Large Scale Heterogeneous IoT Networks," *Computers & Security*, p. 103299, 2023.
- [49] 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.
- [50] J. Roldán, J. Boubeta-Puig, J. L. Martínez, and G. Ortiz, "Integrating complex event processing and machine learning: An intelligent architecture for detecting IoT security attacks," *Expert Systems with Applications*, vol. 149, p. 113251, 2020.
- [51] V. Mothukuri, P. Khare, R. M. Parizi, S. Pouriye, A. Dehghantaha, and G. Srivastava, "Federated-learning-based anomaly detection for iot security attacks," *IEEE Internet of Things Journal*, vol. 9, no. 4, pp. 2545-2554, 2021.
- [52] Y. Yue, S. Li, P. Legg, and F. Li, "Deep learning-based security behaviour analysis in IoT environments: A survey," *Security and Communication Networks*, vol. 2021, pp. 1-13, 2021.
- [53] I. Razzak, K. Zafar, M. Imran, and G. Xu, "Randomized nonlinear one-class support vector machines with bounded loss function to detect of outliers for large scale IoT data," *Future Generation Computer Systems*, vol. 112, pp. 715-723, 2020.
- [54] T. Ergen and S. S. Kozat, "A novel distributed anomaly detection algorithm based on support vector machines," *Digital Signal Processing*, vol. 99, p. 102657, 2020.
- [55] C. Ioannou and V. Vassiliou, "Network attack classification in IoT using support vector machines," *Journal of Sensor and Actuator Networks*, vol. 10, no. 3, p. 58, 2021.
- [56] J. Lesouple, C. Baudoin, M. Spigai, and J.-Y. Tourneret, "How to introduce expert feedback in one-class support vector machines for anomaly detection?," *Signal Processing*, vol. 188, p. 108197, 2021.
- [57] A. Yahyaoui, T. Abdellatif, and R. Attia, "Hierarchical anomaly based intrusion detection and localization in IoT," in *2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)*, 2019: IEEE, pp. 108-113.
- [58] M. Hasan, M. M. Islam, M. I. I. Zarif, and M. Hashem, "Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches," *Internet of Things*, vol. 7, p. 100059, 2019.
- [59] A. P. Agrawal and N. Singh, "Comparative analysis of SVM kernels and parameters for efficient anomaly detection in IoT," in *2021 5th International Conference on Information Systems and Computer Networks (ISCON)*, 2021: IEEE, pp. 1-6.
- [60] Y.-L. Tsou, H.-M. Chu, C. Li, and S.-W. Yang, "Robust distributed anomaly detection using optimal weighted one-class random forests," in *2018 IEEE International Conference on Data Mining (ICDM)*, 2018: IEEE, pp. 1272-1277.
- [61] S. H. Khan, A. R. Arko, and A. Chakrabarty, "Anomaly Detection in IoT Using Machine Learning," in *Artificial Intelligence for Cloud and Edge Computing*: Springer, 2021, pp. 237-254.
- [62] R. Primartha and B. A. Tama, "Anomaly detection using random forest: A performance revisited," in *2017 International conference on data and software engineering (ICoDSE)*, 2017: IEEE, pp. 1-6.
- [63] H. Yang, S. Liang, J. Ni, H. Li, and X. S. Shen, "Secure and efficient knn classification for industrial internet of things," *IEEE Internet of Things Journal*, vol. 7, no. 11, pp. 10945-10954, 2020.
- [64] U. Garg, H. Sivaraman, A. Bamola, and P. Kumari, "To Evaluate and Analyze the Performance of Anomaly Detection in Cloud of Things," in *2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2022: IEEE, pp. 1-7.
- [65] G. E. Selim, E. E. D. Hemdan, A. M. Shehata, and N. A. El - Fishawy, "An efficient machine learning model for malicious activities recognition in water - based industrial internet of things," *Security and Privacy*, vol. 4, no. 3, p. e154, 2021.
- [66] S. Narayanan and S. Uludag, "Two-Tier Anomaly Detection for an Internet of Things Network," in *2023 IEEE 20th Consumer Communications & Networking Conference (CCNC)*, 2023: IEEE, pp. 325-328.
- [67] I. Ullah and Q. H. Mahmoud, "Design and development of RNN anomaly detection model for IoT networks," *IEEE Access*, vol. 10, pp. 62722-62750, 2022.
- [68] Y. Wu, H.-N. Dai, and H. Tang, "Graph neural networks for anomaly detection in industrial internet of things," *IEEE Internet of Things Journal*, vol. 9, no. 12, pp. 9214-9231, 2021.
- [69] M. Saharkhizan, A. Azmoodeh, A. Dehghantaha, K.-K. R. Choo, and R. M. Parizi, "An ensemble of deep recurrent neural networks for detecting IoT cyber attacks using network traffic," *IEEE Internet of Things Journal*, vol. 7, no. 9, pp. 8852-8859, 2020.
- [70] S. Saurav et al., "Online anomaly detection with concept drift adaptation using recurrent neural networks," in *Proceedings of the acm india joint*

- international conference on data science and management of data, 2018, pp. 78-87.
- [71] Y. Wang, M. Perry, D. Whitlock, and J. W. Sutherland, "Detecting anomalies in time series data from a manufacturing system using recurrent neural networks," *Journal of Manufacturing Systems*, vol. 62, pp. 823-834, 2022.
- [72] B. Roy and H. Cheung, "A deep learning approach for intrusion detection in internet of things using bi-directional long short-term memory recurrent neural network," in 2018 28th international telecommunication networks and applications conference (ITNAC), 2018: IEEE, pp. 1-6.
- [73] D. Gaifulina and I. Kotenko, "Selection of deep neural network models for IoT anomaly detection experiments," in 2021 29th Euromicro International Conference on Parallel, Distributed and Network-Based Processing (PDP), 2021: IEEE, pp. 260-265.
- [74] K. Ahmadi and R. Javidan, "Trust Based IOT Routing Attacks Detection Using Recurrent Neural Networks," in 2022 Sixth International Conference on Smart Cities, Internet of Things and Applications (SCIoT), 2022: IEEE, pp. 1-7.
- [75] R. Xu, Y. Cheng, Z. Liu, Y. Xie, and Y. Yang, "Improved Long Short-Term Memory based anomaly detection with concept drift adaptive method for supporting IoT services," *Future Generation Computer Systems*, vol. 112, pp. 228-242, 2020.
- [76] N. Ding, H. Ma, H. Gao, Y. Ma, and G. Tan, "Real-time anomaly detection based on long short-Term memory and Gaussian Mixture Model," *Computers & Electrical Engineering*, vol. 79, p. 106458, 2019.
- [77] X. Zhou, Y. Hu, W. Liang, J. Ma, and Q. Jin, "Variational LSTM enhanced anomaly detection for industrial big data," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 5, pp. 3469-3477, 2020.
- [78] I. Ullah and Q. H. Mahmoud, "Design and development of a deep learning-based model for anomaly detection in IoT networks," *IEEE Access*, vol. 9, pp. 103906-103926, 2021.
- [79] H. Asgharzadeh, A. Ghaffari, M. Masdari, and F. S. Gharehchopogh, "Anomaly-based Intrusion Detection System in the Internet of Things using a Convolutional Neural Network and Multi-Objective Enhanced Capuchin Search Algorithm," *Journal of Parallel and Distributed Computing*, 2023.
- [80] A. Mellit, M. Benghanem, O. Herrak, and A. Messalaoui, "Design of a novel remote monitoring system for smart greenhouses using the internet of things and deep convolutional neural networks," *Energies*, vol. 14, no. 16, p. 5045, 2021.
- [81] N. A. Bajao and J.-a. Sarucam, "Threats Detection in the Internet of Things Using Convolutional neural networks, long short-term memory, and gated recurrent units," *Mesopotamian journal of cybersecurity*, vol. 2023, pp. 22-29, 2023.
- [82] P. Kamat and R. Sugandhi, "Anomaly detection for predictive maintenance in industry 4.0-A survey," in *E3S web of conferences*, 2020, vol. 170: EDP Sciences, p. 02007.
- [83] S. K. Bose, B. Kar, M. Roy, P. K. Gopalakrishnan, and A. Basu, "ADEPOS: Anomaly detection based power saving for predictive maintenance using edge computing," in *Proceedings of the 24th asia and south pacific design automation conference*, 2019, pp. 597-602.
- [84] E. Gultekin and M. S. Aktas, "A Business Workflow Architecture for Predictive Maintenance using Real-Time Anomaly Prediction On Streaming IoT Data," in 2022 IEEE International Conference on Big Data (Big Data), 2022: IEEE, pp. 4568-4575.
- [85] A. Chehri and G. Jeon, "The industrial internet of things: examining how the IIoT will improve the predictive maintenance," in *Innovation in Medicine and Healthcare Systems, and Multimedia: Proceedings of KES-InMed-19 and KES-IIMSS-19 Conferences*, 2019: Springer, pp. 517-527.
- [86] M. Yamauchi, Y. Ohsita, M. Murata, K. Ueda, and Y. Kato, "Anomaly detection in smart home operation from user behaviors and home conditions," *IEEE Transactions on Consumer Electronics*, vol. 66, no. 2, pp. 183-192, 2020.
- [87] A. Lara, V. Mayor, R. Estepa, A. Estepa, and J. E. Díaz-Verdejo, "Smart home anomaly-based IDS: Architecture proposal and case study," *Internet of Things*, vol. 22, p. 100773, 2023.
- [88] S. Ramapatruni, S. N. Narayanan, S. Mittal, A. Joshi, and K. Joshi, "Anomaly detection models for smart home security," in 2019 IEEE 5th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS), 2019: IEEE, pp. 19-24.
- [89] X. Dai, J. Mao, J. Li, Q. Lin, and J. Liu, "HomeGuardian: Detecting Anomaly Events in Smart Home Systems," *Wireless Communications and Mobile Computing*, vol. 2022, 2022.
- [90] S. Hadjixenophontos, A. M. Mandalari, Y. Zhao, and H. Haddadi, "PRISM: Privacy Preserving Internet of Things Security Management," *arXiv preprint arXiv:2212.14736*, 2022.
- [91] C.-R. Su, J. Hajiyev, C. J. Fu, K.-C. Kao, C.-H. Chang, and C.-T. Chang, "A novel framework for a remote patient monitoring (RPM) system with abnormality detection," *Health Policy and Technology*, vol. 8, no. 2, pp. 157-170, 2019.
- [92] M. L. Sahu, M. Atulkar, M. K. Ahirwal, and A. Ahamad, "Cloud-based remote patient monitoring system with abnormality detection and alert notification," *Mobile Networks and Applications*, vol. 27, no. 5, pp. 1894-1909, 2022.
- [93] D. Gupta, M. Gupta, S. Bhatt, and A. S. Tosun, "Detecting anomalous user behavior in remote patient monitoring," in 2021 IEEE 22nd International Conference on Information Reuse and Integration for Data Science (IRI), 2021: IEEE, pp. 33-40.



# Segmentation of Motion Objects in Video Frames using Deep Learning

Feng JIANG<sup>1</sup>, Jiao LIU<sup>2</sup>, Jiya TIAN<sup>3\*</sup>

School of Electrical Information, Changchun Guanghua University, Changchun 130033, China<sup>1</sup>

Student Affairs Office, Changchun Guanghua University, Changchun 130033, China<sup>2</sup>

School of Information Engineering, Xinjiang Institute of Technology, Aksu 843100, China<sup>3</sup>

**Abstract**—The segmentation of the moving objects in the video sequences is one of the most usable series in the machine vision field, which has absorbed the consideration of researchers in the latter decades. It is a challenging task, especially when there are several motion objects in the video, and then the system needs to discover the objects that should be segmented among the trail. Therefore, in this article, we present a new method to segment several motion objects at the same time. In this work, the propagation of the credence of the confidently-estimated frames by fine-tuning the DCNN model with the other frames is the main idea. We exert a DCNN model (which is pre-trained) for the frames to estimate the class of the object; then, we gather the frames where the approximation is locally or globally reliable. In the following, we apply a collection of the frames of CE as the training set to fine-tune the pre-trained network with the existing examples in a video. Our proposed model provides acceptable results, which are better than the results of similar models. These comparisons are made in the dataset of YouTube-VOS. Also, our presented approach is applied in the dataset of DAVIS-2017 and the obtained results are better than the results of the similar works.

**Keywords**—Segmentation; video processing; motion objects; deep convolutional neural network (DCNN)

## I. INTRODUCTION

With the growth of technology and the presence of machines in human life, the various applications of this technology have increased daily. Currently, with the increase in computing capabilities along with the low price of the cameras, image perception is an important part of many applications. One of these applications in image processing is the segmentation of motion objects. The segmentation of the object is the separation of the background and the objects in a trail of video images with a specific purpose [1]. The discovery and segmentation of the moving objects on the trail of videos is a prerequisite step for the high-level systems of machine vision, such as stewardship systems, robotics, and so on. The accuracy of the mentioned systems depends on the segmentation method used. For example, in a surveillance system that uses the information of the movement model for the recognition of people, it should be possible to continuously segment and track the moving objects with high accuracy through the installed cameras in the desired location. Then, by analyzing the received information about the movement and the location of these people, in case of unfortunate events such as falling down which occur, the system can be notified automatically to the relevant centers such as the emergency [2].

With the consideration of the important mentioned applications in the above and many other applications of object segmentation, in the current article, we propose a novel approach for the segmentation of the object in video trials.

Therefore, the main purpose of the segmentation of the video is to separate the foreground from the background with respect to a video trail [3]. Recently, new approaches have been proposed to segment all motion objects in a video and produce larger datasets. This work leads to more challenging tasks [4]. Most of the presented methods in this field evaluate the frames separately [5], and they do not remark on the dimension of the temporal to obtain the affiliation among the successive frames. Recently, an encoder-decoder architecture has been presented based on RNN [6] and is similar to our proposed method.

Therefore, in this paper, the key idea is the propagation of the CE frame credence into another frame using the fine-tuning of the model of DCNN. So, we exert the DCNN model (which is pre-trained) for the frames to estimate the class of the object, and then, we gather the frames where the estimation as globally or locally is reliable. In the following, we exert a collection of the frames of CE as the training set to fine-tune the used pre-trained model with the examples in the videos. Also, we confine the used model of DCNN [7] to only the video. For example, we perform the model centralization in the particular examples in the input video. We, in this procedure, only use the CE region labels and permit the CE frames to determine the un-estimated regions. In addition, we use the feeble labels to prevent the degradation of the model by a few incorrect labels. Our procedures for the generation of the self-consistent datasets and the use of the CE frames for the updation of the system can retrieve the unspecified parts or the classified sections from the frames of UE, which contain several objects.

The article continuation is as the below: Section II characterizes the related works and their overview. Section III characterizes the details of our presented method. The evaluation details and the details of the performed tests are provided in Section IV. In this section also, we provide the visual outcomes and the numerical outcomes of the done tests. In Section V, we provide the suggestions and conclusions.

## II. RELATED WORKS

Due to the wide applications of the segmentation of motion objects in the arena of machine vision, researchers have studied and have presented different methods for this task in recent

years. Among the comprehensive performed works in this field is the presented work in [8], which has reviewed and classified the proposed segmentation methods. In [9]–[12], the different methods for the segmentation and the tracking of the motion objects have been investigated. Usually, the segmentation is done based on the obtained information for a series of special characteristics from the objects. These characteristics include the below cases: the edge, the texture, the color of the objects, the movement information, the corner points, the appearance of the objects, etc. Any segmentation algorithm based on the application can use any of these characteristics or a combination of these characteristics. For example, in the algorithms that segment and track the objects based on the object contour, the edge feature [13] is used. In [14], the motion objects were segmented based on the difference between the existing edges between two consecutive frames. The detectors of the corner points in the literature on object segmentation are Moravec [15], Harris [16], KLT [17], and SIFT [18].

In addition, the segmentation of moving objects based on deep learning techniques has received regard in the association of the research in the latter years. It can be due to the emergence of novel segmentation datasets and new challenges: Berkeley (2011), SegTrack (2013) [19], Berkeley Freiburg (2014) [20], DAVIS (2016-2017) [21], and YouTubeVOS (2018) [22]. These datasets provide the biggest content of the tagged videos.

The later works, such as [15], use the optical flow for the temporal adaptation after the use of the fields of Markov random, which is the basis on the taken specifications of a CNN model. The other suggestion for the obtention of the coherence of the temporal is the use of the boded masks on prior frames as a guide for the subsequent frames [7]. The proposed method in [23] disseminates the information using spatiotemporal features. Finally, the proposed method in [24] uses an architecture of the encoder-decoder RNN that employs the LSTM for the learning of the trail.

In the segmentation of the objects in the video, the learning with the single-shot is found as the use from an alone tagged frame for the estimation of the residual frame's segmentation in a sequence. Also, the learning with the zero-shot is found as the construction models, which do not require the initialization for the generation of the masks of the segmentation of the object in the trail of the video. There are multiple articles in the literature which is emphasized the first mask for the input to can propagate via the trail [3], [7], [10], [25], and [26]. Generally, the approaches with the single-shot outperform in comparison to the approaches with the zero-shot because the first segmentation is formerly taken, so there is no need for the estimation of the mask of the first segmentation of the

abrasion. Most of the proposed systems emphasize online learning, which is the adaption of the weights with the first frame and associated masks. Usually, the methods of online learning achieve better outcomes, but they need more computing time. On the learning with the zero-shot, for the estimation of the segmentation of the object on a video, multiple papers have used the object saliency [8], [27], [28] or they have used the object suggestion methods outputs [12], or they have used network with two-stream. The exploitation of the motion templates on the videos is perused at [29], but the article of [14] formulates the 3D representation conclusion of a planar object and the motion segmentation. In addition, foreground segmentation which is the basis of the sample embedding is presented in [16].

Also, optical flow computation is one of the fundamental tasks in computer vision. Deep learning methods allow efficient computation of optical flow, both in supervised learning on synthetic data [42], and in the self-supervised [39] setting. Additionally, in [40], the authors propose to highlight the independently moving object by compensating for the background motion, either by registering consecutive frames, or explicitly estimating camera motion. Another line of work has tackled the problem by explicitly leveraging the independence, in the flow field, between the moving object and its background. For instance, [41] proposes an adversarial setting, where a generator is trained to produce masks, altering the input flow, such that the inpainter fails to estimate the missing information.

Finally, in [43-45], two protocols have attracted increasing interest from the vision community, namely, semi-supervised video object segmentation (semi-supervised VOS), and unsupervised video object segmentation (unsupervised VOS). The former aims to re-localize one or multiple targets that are specified in the first frame of a video with pixel-wise masks, and the latter considers automatically separating the object of interest (usually the most salient one) from the background in a video sequence.

### III. PROPOSED METHOD

To present our method, we consider an important hypothesis. We presume that the video contains at least some frames of CE such that it is useful for the improvement of the uncertainly-estimated frame outcomes. In the presented method, the main idea is the propagation of the credence of the frames of CE using the fine-tuning of DCNN. Therefore, our method includes the below stages: the election of the CE frames, the production of the label mapping, and the matching of a model with the input video. In the following subsections, we characterize the desired algorithm of these stages. Fig. 1 displays the general format of our presented approach.

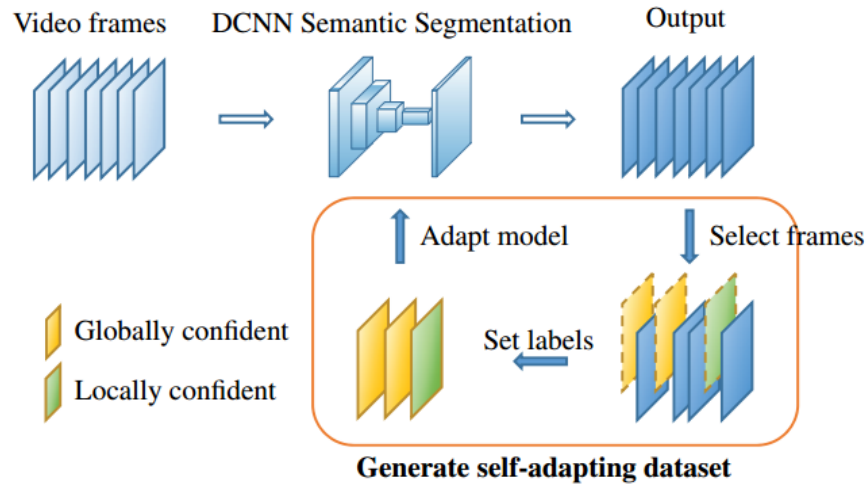


Fig. 1. The general format of our presented approach.

### A. Steps of Our Presented Method

Fig. 1 displays the general steps of our presented method. In this sub-section, we will characterize the presented approach details to segment the motion objects in the video frames.  $F$  shows the collection of the indices of the frames; also  $W$  displays the collection of the desired poor labels from the input video. The presented method starts using the DCNN model  $\theta$  which is pre-trained (for the frame  $f \in F$ ), then we apply *Softmax* for the computation of the probability  $P(x_i|\theta)$  where  $i$ -th pixel is the organ of the class  $x_i \in O$  which in it,  $O$  represents a collection of the classes of the object and the classes of the background. The mapping of the semantic label  $S$  can be measured by the use of the *argmax* for each pixel  $i$ :

$$S(i) = \arg \max_{x_i} P(x_i|\theta) \quad (1)$$

We gather the self-adaptive dataset  $G$  for adaption of the model of DCNN with the input video, which  $G$  includes the frames of CE and the related tag mappings. We gather the global CE frames and the local CE frames. Then we calculate corresponding label mappings  $G^g$  and  $G^l$  for the make of a self-consistent collection. Algorithm 1 summarizes the processes of the frame selection and the label calculation. First, we apply the analysis of the connected part in each mapping from the class  $S$  for the generation of the collection of the candidate object zones. For  $k$ -th mapping of  $R_k \in R$ , the confidence of  $\mathcal{C}(R_k)$  measures the evaluated zones, which in it, the operator  $\mathcal{C}(\cdot)$  catches the label mapping as the input. Then, it calculates the mean probability of which pixels are labeled as objects. The label mapping has the labels of the related class.

In the following, we construct the mapping of the label  $G_f^g$  with the setting of the zone label when the confidence value

trespasses an upper threshold  $t_0$ . Also, we adjust the label of the background for each pixel where  $P(x_i = bg|\theta)$  (for being the background) is the greater than the threshold of  $t_b$ .

For the completion of  $G_f^g$ , the residual undefined zones must be processed. We, for this goal, let the residual pixels be labeled as "ignored." The pixels of the unspecified "ignored" are not attended to in the calculation of the loss value for the updation of the model. Also, we relinquish all pixels which have tags that are not on the collection of  $W$ . We surcharge the global frames of CE with  $G_f^g$  which have one safe zone for self-consistent dataset  $G$ .

Since the elected frames may be distributed temporally, our model can be overcome by the frames which are elected in a short time. For the reduction of the obtained error and for the regularization of the model, it is recommended to select the local frames of CE which have the best confidence of the object in each interval  $\tau b$ . We determine the local CE frames and their label mapping  $G^l$  as follows: For each frame  $f$ , we create a label mapping  $G_f^l$  using label keeping of total pixels if and only if  $S(i)$  to be consisted on  $W$ , when we set the background as the prior. In the following, we compute the frame confidence by the computation of  $\mathcal{C}(G_f^l)$  and then we consider the frame with the highest confidence during each part of the frame  $\tau b$  as the local CE. Let the local frame of CE formerly not elected as the global frame of CE; then we surcharge it into the self-adaptive dataset  $G$ .

With the consideration of the self-adaptive dataset  $G$  which is computed by the mentioned processes, finally, we reconcile the model  $\theta$  with the video. This task is done using the fine-tuning of the model into  $\theta'$ . In the following, we calculate the novel label mapping using  $\theta'$  for each frame.

<b>Algorithm 1. Procedures for Selecting the Frames and Calculating the Labels</b>	
Input: DCNN model $\theta$ , a set of weak labels $W$	
Local best confidence $d = 0$	
<b>for</b> $f \in F$ <b>do</b>	
Initialize $G_f^g, G_f^l$ to ignored label	
Compute $P(x \theta)$ and $S = \mathit{arg\ max}_x P(x \theta)$	
Compute set $R$ of connected components in $S$	
<b>for</b> $R_k \in R$ <b>do</b>	
<b>if</b> $S(i) \notin W, i \in R_k$ <b>then continue</b>	
<b>if</b> $C(R_k) > t_0$ <b>then</b>	
Set $G_f^g(i) = S(i), \forall i \in R_k$	
Set $G_f^l(i) = S(i), \forall i \in R_k$	
Set $G_f^g(i) = G_f^l(i) = 0, \forall i \text{ st. } P(x_i = \mathit{bg} \theta) > t_b$	
<b>if</b> $C(G_f^g) > 0$ <b>then</b> $G \leftarrow G \cup \{G_f^g\}$	
<b>if</b> $C(G_f^l) > d$ <b>then</b>	
Update $t = f$ and $d = C(G_f^l)$	
<b>if</b> $f \bmod \tau b = 0$ <b>then</b>	
<b>if</b> $G_f^g \notin G$ <b>then</b>	
	$G \leftarrow G \cup \{G_f^l\}$
Initialize $d = 0$	
Fine-tune DCNN model $\theta$ to $\theta'$ using the set $G$	

### B. Development of Proposed Method for Un-Supervised Video

Our presented approach can be used for the processing of unsupervised video. This task can easily be applied to unsupervised videos using the limitation of line eight in Algorithm 1. This omission means that the model doesn't manage whether the class emerges really on the video. So, we adjust all tags of the CE zones even if the tags are wrong. The experiments show that most processed videos have the same outcomes as the weakly-supervised videos is so much that the pixel tags specified by a great probability typically match the true tags. However, exceptions occur, which these exceptions are related to incorrect labels. They can degrade the model, and they can reduce the accuracy compared to the settings of weakly supervised.

### C. Implement the Post-Processing for the Correction

Since the DCNN output is not sufficient to accurately characterize the object therefore, we apply the fully-connected CRF [30]. We apply the DCNN output for the single expression. Also, we apply the pixel's positions and the pixels' colors for the calculation of the even expressions (similar to [31]). We, finally, modify the label mapping via the practices of morphology (such as erosion and dilation).

## IV. TESTS AND THE OUTCOMES EVALUATION

In this sub-section, first, we will present the implementation details and the performed tests. Also, we will introduce the used dataset. In the following, the tests' results are presented, and an analytical evaluation is done.

### A. Details of Implementation

The tests in this article are performed as the single-shot and the zero-shot. Also, the designed tests are done using two datasets: YouTube-VOS [32] and DAVIS-2017 [33]. The first dataset, YouTube-VOS, contains 474 films on the set of the validation and 3471 films on the set of the training. It is the biggest dataset in the field of the segmentation of the video object. In addition, the training dataset contains 65 unique

groups of the object, which are considered as the observed groups. Also, in the dataset of the validation, there are 91 groups of the object that consist of 26 unseen groups and all seen groups.

On the other hand, the dataset of DAVIS-2017 includes 60 films for the training dataset, 30 films for the validation dataset, and 30 films for the test dataset. In both datasets, the videos contain several objects, and their duration is between three to six seconds. The Python programming language has been used for the implementation of these tests. The presented method is implemented in a machine with Core (TM) i7 CPU 3.0 GHz Intel(R) and 8G RAM. The convolutional network is implemented on GPU, and the used graphic card in this method is NVIDIA GEFORCE 840M. The tests are analyzed by the use of the normal analysis criteria: (1) the accuracy of the contour  $F$  and (2) the similarity of the region  $J$ . On the YouTube-VOS dataset, these criteria are divided into two sub-criteria, depending on whether groups already have been seen with a network ( $F_{seen}$  and  $J_{seen}$ ) or have not been seen by the model ( $J_{unseen}$  and  $F_{unseen}$ ). The concept of the seen (or the unseen) means that, these categories are included in the set of training (or are not included).

### B. Experiments and Results for the YouTube-VOS Dataset

As mentioned, the tests and the results are presented in two modes: the single-shot and the zero-shot. The single-shot mode consists of the object segmentation of a video according to the mask of the objects on the initial frame. But zero-shot mode involves the video objects segmentation without the previous data about that which of the objects must be segmented. It means that no object mask is obtained. This work is more complicated than the single-shot mode because the network must identify and then segment the objects that appeared in the film.

Table I shows the obtained results on the validation dataset of YouTube-VOS for the single-shot mode. All presented models in this study were trained using an 80-20 split for the training dataset. Fig. 2 shows some qualitative results, which in it, we can view, which our proposed method better maintains

the segmentation of the objects over time. The proposed network can learn how the fixing the faults that may arise in the deduction. Table I can view which this approach is strong and has a suitable performance. Fig. 3 displays some qualitative outcomes which compare our trained approach over the mask of the ground truth and our trained approach over the concluded mask.

Table II displays the comparison of our presented model and similar approaches using the entire training dataset of YouTube-VOS. As it is clear, our proposed model has analogous outcomes with the mentioned model in [32]. The proposed method has an awhile worse turnover for the similarity of the region  $J$ . However, it has an awhile better turnover for the accuracy of the contour  $F$ . The proposed network performs better than the remaining advanced methods [25], [34]–[36] for the observed categories. Also, depending on the number of examples in the videos, Table III displays the related outcomes to the similarity of the region  $J$  and the accuracy of the contour  $F$ . We can view which objects for segmentation be fewer, then the work is easier, and we get better outcomes for the trails with only one or two annotated objects. Fig. 4 displays the qualitative outcomes for our presented approach for different trials from the validation set of YouTube-VOS. It contains the samples by the different samples number. Note that which samples are segmented correctly. However, there are different samples of a similar group on the video trail (the leopard, the sheep, the bird, the fish, or the person), or there are cases that vanish from the trail (a sheep on third row and a dog in fourth row).

So far, we have presented the test outcomes for our presented approach in the single-shot mode for the YouTube-VOS dataset. Also, we provide the outcomes for the mentioned dataset on the zero-shot mode. It should be noted that today, there is no designed special dataset for zero-shot segmentation. Although the YouTubeVOS dataset and the DAVIS-2017 dataset can be used to train and evaluate the models without the use of the provided annotations in initial frame, these datasets have this restriction in which the total appeared objects on the film are not annotated. In the YouTube-VOS dataset, specifically, a maximum of five object instances in each video are annotated. This makes sense when the objects are given for the segmentation but may be problematic for the zero-shot segmentation because the model can correctly segment the objects which are not annotated on the dataset. Fig. 5 displays some samples in it and some annotations of the missing object. Notwithstanding the stated quandary on the annotations of the missing object, for this mode, we have trained our network using the existing object annotations in this dataset.

Table IV displays the obtained outcomes for the validation dataset of YouTube-VOS on the segmentation with the zero-shot. Similar to the segmentation problem in single-shot mode, the proposed model has a good performance for object segmentation in the video. Fig. 6 displays some outcomes for segmentation in the zero-shot mode on the validation dataset of YouTube-VOS. Note that the masks are not obtained, so the system must detect the objects that must be segmented.

### C. Experiments and Outcomes for the DAVIS-2017 Dataset

We test our pre-trained model (which is trained using the YouTube-VOS dataset) on a different dataset: DAVIS-2017. As shown in Table V, if the pre-trained network is done directly for the DAVIS-2017 dataset, our presented approach performs better than the rest of the approaches that do not use online learning. In addition, when the proposed model is adjusted for the training dataset of DAVIS-2017, the proposed method outperforms some methods (for example, OSVOS [3]). Fig. 7 displays the obtained visual outcomes on the dataset of DAVIS-2017 in the single-shot mode.

But in the zero-shot mode, by reviewing the articles and the research literature, it was found that there are no formal outcomes for the zero-shot mode on the DAVIS-2017 dataset so that we can compare our proposed method with it. The segmentation with zero-shot mode only is remarked for the DAVIS-2016 dataset, which in the unsupervised approaches have been made for it. Using the YouTube-VOS dataset on the zero-shot mode, if our model is directly done on the DAVIS-2017 dataset, our pre-trained model yields an average region similarity equal to  $J = 22.4$  and an average contour accuracy equal to  $F = 28.0$ . If this pre-trained network is fine-tuned using the validation dataset of DAVIS-2017, then it results in a while better efficiency:  $F = 30.6$  and  $J = 24.7$ . This poor efficiency of the zero-shot segmentation on the DAVIS-2017 dataset can be illustrated by the poor efficiency of the YouTube-VOS dataset in the unseen groups. Fig. 8 displays the visual outcomes of the test dataset of DAVIS-2017, which in it the mask of the object is not obtained.

### D. Discussion

The results obtained from the qualitative experiments showed that our proposed method maintains the segmentation of the objects as better over time. This is because the proposed network can learn how to fix the errors that may occur in the deduction. Also, the results showed that when we test our pre-trained model on a different dataset, our proposed approach outperforms other approaches that do not use the online learning. Furthermore, when the proposed model is adjusted to DAVIS-2017 training dataset, the proposed method outperforms other approaches. Also, the various tests on YouTube-VOS showed that if our model is run directly on the DAVIS-2017 dataset, our pre-trained model yields an average regional similarity of  $J=22.4$  and an average contour accuracy of  $F= 28.0$ . If this pre-trained network is fine-tuned by using the DAVIS-2017 validation dataset, it gives the better performance for some time:  $F=30.6$  and  $J=24.7$ . This poor performance of the zero-shot segmentation on the DAVIS-2017 dataset can be illustrated by the poor performance of the YouTube-VOS dataset on the unseen groups. Finally, the presented method in this article has a specific limitation that occurs sometimes. This limitation occurs when a video does not match the hypothesis that we stated at the beginning (at least one zone of the object collates with the classes of the pre-trained object, or at least one frame has the true tag). Mostly, these samples happen due to the size of very small the objects on a video. The lack of a frame for improvement of other frames gives similar outcomes to the base model. The researchers can remark on this limitation in their subsequent works.

TABLE I. OBTAINED RESULTS FOR VALIDATION DATASET OF YOUTUBE-VOS IN THE SINGLE-SHOT MODE

YouTube-VOS Dataset in the Single-Shot Mode				
	$J_{seen}$	$J_{unseen}$	$F_{seen}$	$F_{unseen}$
Proposed Method	64.0	45.1	67.9	51.2

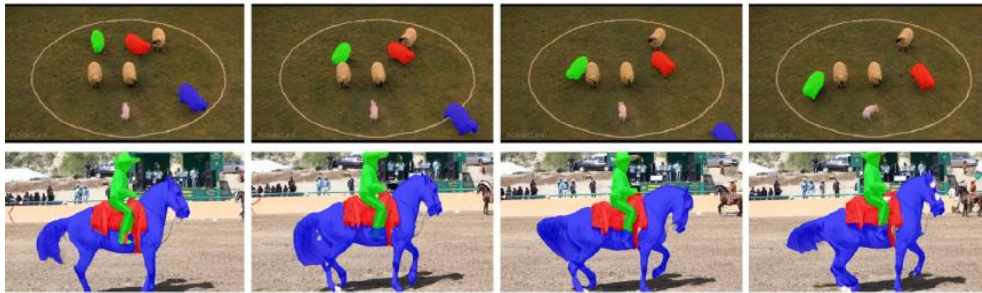


Fig. 2. Qualitative results for our presented approach in a single-shot mode for the YouTube-VOS dataset.

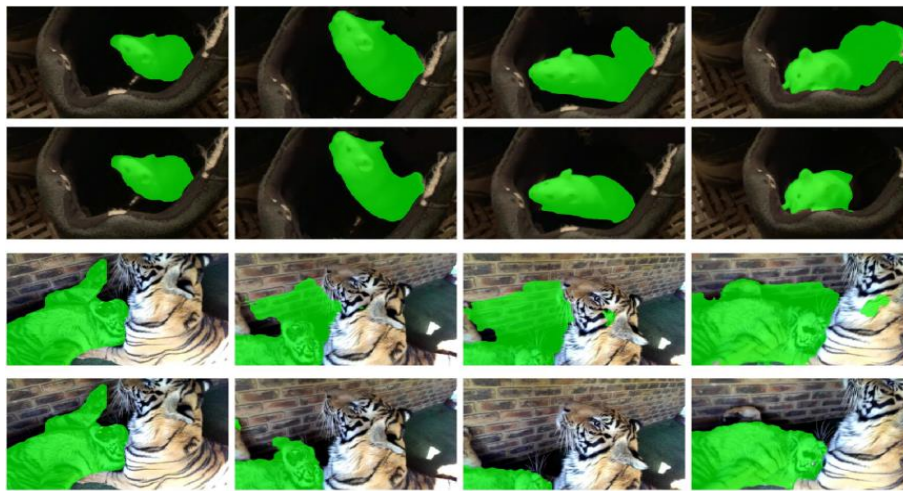


Fig. 3. Some comparative qualitative outcomes for our presented approach on the single-shot mode for the YouTube-VOS dataset.

TABLE II. COMPARISON OF OUR PRESENTED APPROACH WITH THE ADVANCED APPROACHES FOR THE SINGLE-SHOT MODE ON THE VALIDATION SET OF YOUTUBE-VOS. THE TERM OL REFERS TO ONLINE LEARNING

YouTube-VOS Dataset in Single-Shot Mode					
	OL	$J_{seen}$	$J_{unseen}$	$F_{seen}$	$F_{unseen}$
OSVOS [3]	Yes	59.8	54.2	60.5	60.7
MaskTrack [25]	Yes	59.9	45.0	59.5	47.9
OnAVOS [35]	Yes	60.1	46.6	62.7	51.4
OSMN [36]	No	60.0	40.6	60.1	44.0
S2S w/o OL [32]	No	66.7	48.2	65.5	50.3
Proposed Method	No	64.8	45.4	68.4	51.9

TABLE III. ANALYSIS OF OUR PRESENTED APPROACH DEPENDING ON THE NUMBER OF SAMPLES IN THE SEGMENTATION OF THE SINGLE-SHOT

Number of Samples (YouTube-VOS)					
	1	2	3	4	5
$J_{mean}$	78.9	63.3	51.2	50.6	56.9
$F_{mean}$	76.1	68.1	56.4	62.9	66.8

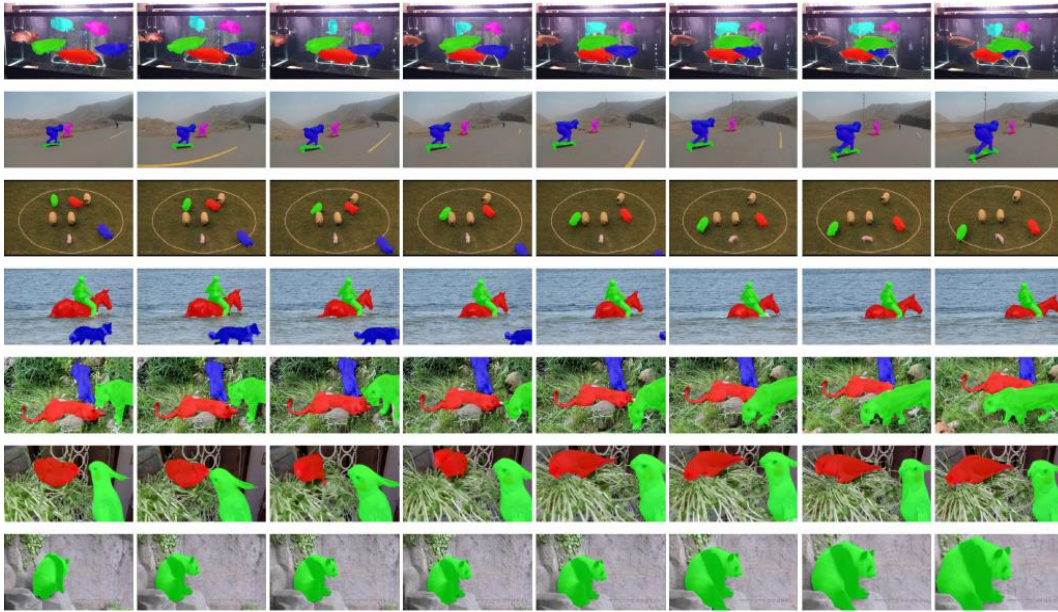


Fig. 4. Some visual results for our presented approach for the different trials from the YouTube-VOS validation set.



Fig. 5. The examples of the missing object annotations.

TABLE IV. THE RESULTS OF THE PERFORMED TESTS FOR THE SEGMENTATION WITH THE ZERO-SHOT MODE FOR THE DATASET OF YOUTUBE-VOS

YouTube-VOS Dataset in the Zero-Shot Mode				
	$J_{seen}$	$J_{unseen}$	$F_{seen}$	$F_{unseen}$
Proposed Method	45.2	24.1	45.9	24.2

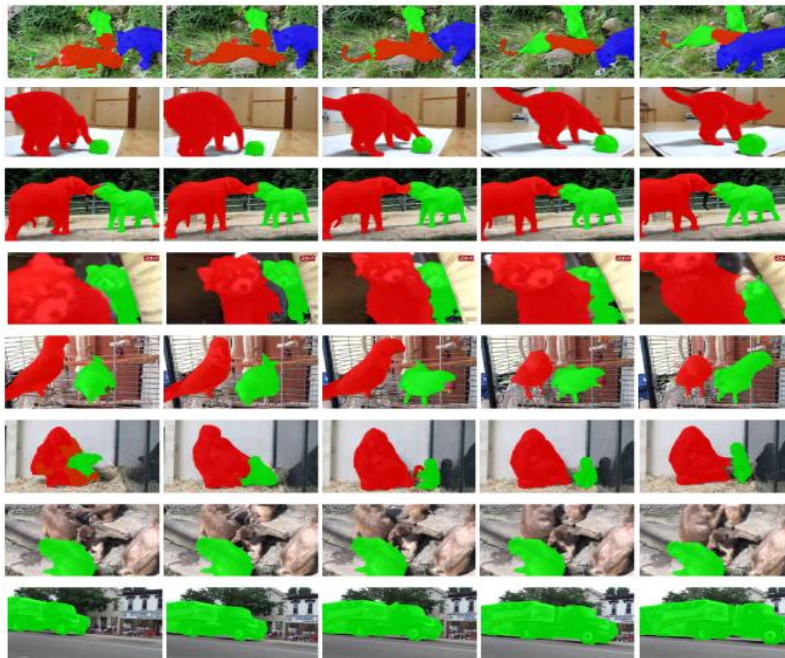


Fig. 6. Qualitative outcomes for segmentation with the zero-shot mode for the dataset of YouTube-VOS.

TABLE V. COMPARISON OF OUR PRESENTED APPROACH OVER ADVANCED METHODS FOR SEGMENTATION IN THE SINGLE-SHOT MODE IN THE DAVIS-2017 DATASET. NOTE THAT OL REFERS TO ONLINE LEARNING

DAVIS-2017 Dataset in the Single-Shot Mode			
	OL	J	F
OSVOS [3]	Yes	47.0	54.8
OSVOS-S [37]	Yes	52.9	62.1
CINM [38]	Yes	64.5	70.5
OnAVOS [35]	Yes	52.9	62.1
OSMN [36]	No	37.7	44.9
FAVOS[4]	No	42.9	44.2
Proposed Method	No	48.8	53.3

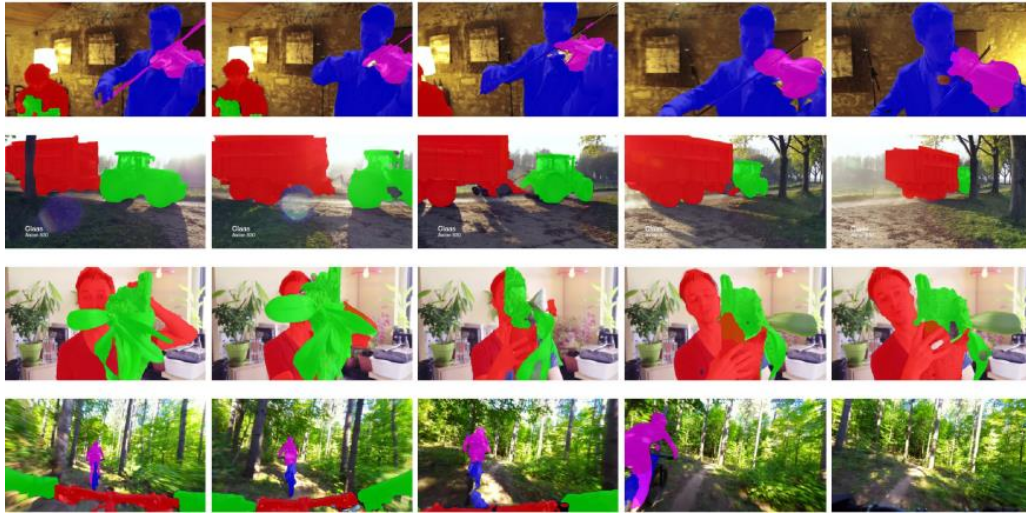


Fig. 7. Visual outcomes of the single-shot segmentation on the dataset of DAVIS-2017.

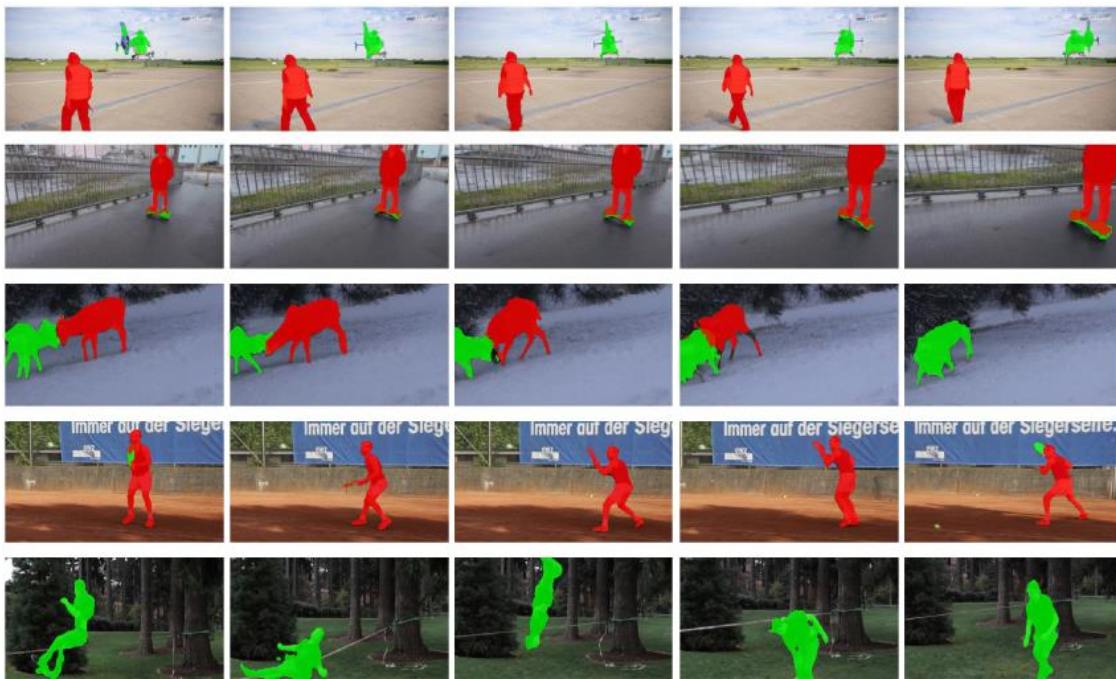


Fig. 8. Visual outcomes of the test dataset of DAVIS-2017 in the zero-shot segmentation mode.



## V. CONCLUSIONS AND SUGGESTIONS

In our article, we propose a novel method for the segmentation of the motion objects which exist in the video. This method reconciles the model of the pre-trained DCNN with the input film. For the fine-tuning of the model, which is trained as vastly to be special for the video, we created a self-adaptive set that includes multiple frames, which helps to the improvement of the results of the frames of UE. This model is designed for segmentation in the single-shot mode and the zero-shot mode. Also, this model is applied in the YouTube-VOS dataset and the DAVIS-2017 dataset. The tests display that our trained model has a better performance than similar methods. In addition, our model improves the performance of similar methods. For future research, it is suggested to develop a semi-supervised film framework for the accuracy increase. It can also be expected that this efficient self-adaptive method can generate video datasets with accurate labels.

## ACKNOWLEDGMENT

This work was supported by “Sponsored by Natural Science Foundation of Xinjiang Uygur Autonomous Region”(2022D01C461), “Research on UAV moving target detection and tracking system based on computer vision”(ZZ202105).

## REFERENCES

- [1] A. Yilmaz, O. Javed, and M. Shah, “Object tracking: A survey,” *Acm computing surveys (CSUR)*, vol. 38, no. 4, pp. 13-es, 2006.
- [2] J. K. Aggarwal and Q. Cai, “Human motion analysis: a review. Nonrigid and Articulated Motion Workshop, 1997,” *Proceedings.*, IEEE, pp. 90–102, 1997.
- [3] S. Caelles, K.-K. Maninis, J. Pont-Tuset, L. Leal-Taixé, D. Cremers, and L. Van Gool, “One-shot video object segmentation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 221–230.
- [4] J. Cheng, Y.-H. Tsai, W.-C. Hung, S. Wang, and M.-H. Yang, “Fast and accurate online video object segmentation via tracking parts,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7415–7424.
- [5] J. Cheng, Y.-H. Tsai, S. Wang, and M.-H. Yang, “Segflow: Joint learning for video object segmentation and optical flow,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 686–695.
- [6] 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*, 2016, pp. 770–778.
- [7] Y.-T. Hu, J.-B. Huang, and A. Schwing, “Maskrcnn: Instance level video object segmentation,” *Adv Neural Inf Process Syst*, vol. 30, 2017.
- [8] H. P. Moravec, “Visual mapping by a robot rover,” in *Proceedings of the 6th international joint conference on Artificial Intelligence-Volume 1*, 1979, pp. 598–600.
- [9] C. Harris and M. Stephens, “A combined corner and edge detector,” in *Alvey vision conference*, Citeseer, 1988, pp. 10–5244.
- [10] J. Shi, “Good features to track,” in *1994 Proceedings of IEEE conference on computer vision and pattern recognition*, IEEE, 1994, pp. 593–600.
- [11] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *Int J Comput Vis*, vol. 60, pp. 91–110, 2004.
- [12] K. Mikolajczyk and C. Schmid, “A performance evaluation of local descriptors,” *IEEE Trans Pattern Anal Mach Intell*, vol. 27, no. 10, pp. 1615–1630, 2005.
- [13] K. Mikolajczyk and C. Schmid, “A performance evaluation of local descriptors,” *IEEE Trans Pattern Anal Mach Intell*, vol. 27, no. 10, pp. 1615–1630, 2005.
- [14] C. Stauffer and W. E. L. Grimson, “Learning patterns of activity using real-time tracking,” *IEEE Trans Pattern Anal Mach Intell*, vol. 22, no. 8, pp. 747–757, 2000.
- [15] R. Zhang and J. Ding, “Object tracking and detecting based on adaptive background subtraction,” *Procedia Eng*, vol. 29, pp. 1351–1355, 2012.
- [16] C. Hua, H. Wu, Q. Chen, and T. Wada, “Object tracking with target and background samples,” *IEICE Trans Inf Syst*, vol. 90, no. 4, pp. 766–774, 2007.
- [17] C. Hua, H. Wu, Q. Chen, and T. Wada, “K-means Tracker: A General Algorithm for Tracking People,” *J. Multim.*, vol. 1, no. 4, pp. 46–53, 2006.
- [18] C. Hua, H. Wu, Q. Chen, and T. Wada, “K-means clustering based pixel-wise object tracking,” *Information and Media Technologies*, vol. 3, no. 4, pp. 820–833, 2008.
- [19] P. Ochs, J. Malik, and T. Brox, “Segmentation of moving objects by long term video analysis,” *IEEE Trans Pattern Anal Mach Intell*, vol. 36, no. 6, pp. 1187–1200, 2013.
- [20] F. Perazzi, A. Khoreva, R. Benenson, B. Schiele, and A. Sorkine-Hornung, “Learning video object segmentation from static images,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2663–2672.
- [21] 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 *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 724–732.
- [22] J. Pont-Tuset, F. Perazzi, S. Caelles, P. Arbeláez, A. Sorkine-Hornung, and L. Van Gool, “The 2017 davis challenge on video object segmentation,” *arXiv preprint arXiv:1704.00675*, 2017.
- [23] M. Ren and R. S. Zemel, “End-to-end instance segmentation with recurrent attention,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 6656–6664.
- [24] B. Romera-Paredes and P. H. S. Torr, “Recurrent instance segmentation,” in *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VI 14*, Springer, 2016, pp. 312–329.
- [25] F. Perazzi, A. Khoreva, R. Benenson, B. Schiele, and A. Sorkine-Hornung, “Learning video object segmentation from static images,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2663–2672.
- [26] O. Russakovsky et al., “Imagenet large scale visual recognition challenge,” *Int J Comput Vis*, vol. 115, pp. 211–252, 2015.
- [27] A. Salvador et al., “Recurrent neural networks for semantic instance segmentation,” *arXiv preprint arXiv:1712.00617*, 2017.
- [28] H. Song, W. Wang, S. Zhao, J. Shen, and K.-M. Lam, “Pyramid dilated deeper convlstm for video salient object detection,” in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 715–731.
- [29] P. Tokmakov, K. Alahari, and C. Schmid, “Learning motion patterns in videos,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3386–3394.
- [30] P. Krähenbühl and V. Koltun, “Efficient inference in fully connected crfs with gaussian edge potentials,” *Adv Neural Inf Process Syst*, vol. 24, 2011.
- [31] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, “Semantic image segmentation with deep convolutional nets and fully connected crfs,” *arXiv preprint arXiv:1412.7062*, 2014.
- [32] N. Xu et al., “Youtube-vos: A large-scale video object segmentation benchmark,” *arXiv preprint arXiv:1809.03327*, 2018.
- [33] J. Pont-Tuset, F. Perazzi, S. Caelles, P. Arbeláez, A. Sorkine-Hornung, and L. Van Gool, “The 2017 davis challenge on video object segmentation,” *arXiv preprint arXiv:1704.00675*, 2017.
- [34] S. Caelles, K.-K. Maninis, J. Pont-Tuset, L. Leal-Taixé, D. Cremers, and L. Van Gool, “One-shot video object segmentation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 221–230.
- [35] P. Voigtlaender and B. Leibe, “Online adaptation of convolutional neural networks for video object segmentation,” *arXiv preprint arXiv:1706.09364*, 2017.

- [36] L. Yang, Y. Wang, X. Xiong, J. Yang, and A. K. Katsaggelos, "Efficient video object segmentation via network modulation," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 6499–6507.
- [37] K.-K. Maninis et al., "Video object segmentation without temporal information," IEEE Trans Pattern Anal Mach Intell, vol. 41, no. 6, pp. 1515–1530, 2018.
- [38] L. Bao, B. Wu, and W. Liu, "CNN in MRF: Video object segmentation via inference in a CNN-based higher-order spatio-temporal MRF," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 5977–5986.
- [39] Liang Liu, Jiangning Zhang, Ruifei He, Yong Liu, Yabiao Wang, Ying Tai, Donghao Luo, Chengjie Wang, Jilin Li, and Feiyue Huang. Learning by analogy: Reliable supervision from transformations for unsupervised optical flow estimation. In Proc. CVPR, 2020.
- [40] Hala Lamdouar, Charig Yang, Weidi Xie, and Andrew Zisserman. Betrayed by motion: Camouflaged object discovery via motion segmentation. Proc. ACCV, 2020.
- [41] Tianfei Zhou, Shunzhou Wang, Yi Zhou, Yazhou Yao, Jianwu Li, and Ling Shao. Motion-attentive transition for zero-shot video object segmentation. In AAAI, volume 34, pages 13066–13073, 2020.
- [42] Zachary Teed and Jia Deng. Raft: Recurrent all-pairs field transforms for optical flow. In Proc. ECCV, 2020.
- [43] Pavel Tokmakov, Cordelia Schmid, and Karteek Alahari. Learning to segment moving objects. International Journal of Computer Vision, 2019.
- [44] Paul Voigtlaender, Yuning Chai, Florian Schroff, Hartwig Adam, Bastian Leibe, and Liang-Chieh Chen. Feelvos: Fast end-to-end embedding learning for video object segmentation. In Proc. CVPR, 2019.
- [45] Wenguan Wang, Xiankai Lu, Jianbing Shen, David J Crandall, and Ling Shao. Zero-shot video object segmentation via attentive graph neural networks. In Proc. ICCV, 2019.

# Classification of Coherence Indices Extracted from EEG Signals of Mild and Severe Autism

Lingyun Wu

School of Culture Communication, Henan Vocational Institute of Arts, Zhengzhou 451464, Henan, China

**Abstract**—Autism spectrum disorder is a debilitating neurodevelopmental illness characterized by serious impairments in communication and social skills. Due to the increasing prevalence of autism worldwide, the development of a new diagnostic approach for autism spectrum disorder is of great importance. Also, diagnosing the severity of autism is very important for clinicians in the treatment process. Therefore, in this study, we intend to classify the electroencephalogram (EEG) signals of mild and severe autism patients. Twelve patients with mild autism and twelve patients with severe autism with the age range of 10-30 years participated in the present research. Due to the difficulties of working with autism patients and recording EEG signals from these patients in the awake state, the Emotiv Epoch headset device was utilized in this work. After signal preprocessing, we calculated short-range and long-range coherence values in the frequency range of 1-45 Hz, including short- and long-range intra- and inter-hemispheric coherence features. Then, statistical analysis was conducted to select coherence features with statistical differences between the two groups. Multilayer perceptron (MLP) neural network and support vector machine (SVM) with radial basis function (RBF) kernel were used in the classification stage. Our results showed that the best MLP classification performance was obtained by selected inter-hemispheric coherence features with accuracy, sensitivity and specificity of 96.82%, 97.82% and 96.92%, respectively. Also, the best SVM classification performance was obtained by selected inter-hemispheric coherence features with accuracy, sensitivity and specificity of 94.70%, 93.85% and 95.55%, respectively. However, it should be noted that the MLP neural network imposes a much higher computational cost than the SVM classifier. Considering that our simple system gives promising results in diagnosing autistic patients with mild and severe severities from EEG, there is scope for further work with a larger sample size and different ages and genders.

**Keywords**—Autism spectrum disorder; electroencephalography (EEG); classification; neural network; support vector machine; coherence feature

## I. INTRODUCTION

Autism spectrum disorder is a debilitating neurodevelopmental illness characterized by serious impairments in communication and social skills. Patients with autism manifest repetitive and restricted behavior [1-3]. The prevalence of autism has increased considerably from 0.67% in 2000 to 2.58% in 2016 in the United States [4-6]. Therefore, in recent years, researchers emphasized different approaches to early diagnosis of this disorder to provide timely intervention and achieve better treatment outcomes [7, 8]. In this regard, it is also very important to determine the severity of the disease in autism spectrum disorder because it leads to changing

treatment approaches due to the level of brain disorders [9]. Brain images have shown abnormalities in brain and head size and limbic and cerebellar structure in autistic patients [10]. Magnetic resonance imaging (MRI) studies have shown that autistic patients have serious disturbances in the functional connectivity between different brain regions [11, 12]. fMRI and diffusion tensor imaging (DTI) studies also showed that patients with autism had reduced long-range functional connectivity at rest and during different executive cognitive tasks [13]. However, due to its high availability, cheapness, and ease of use, many researchers have analyzed the electroencephalogram (EEG) signals of autism patients to study their brain function [14-18]. For instance, in a recent study, Hadoush et al. [19] analyzed and compared the brain complexity of children with mild and severe autism through multiscale entropy analysis of EEG signals. They found that the brain complexity of children with mild autism is higher than that of children with severe autism in the right parietal, right frontal, left parietal and central cortices. Finally, they concluded that EEG multiscale entropy could serve as a sensitive index to detect the level of autism severity. In a review study, Wang et al. suggested excessive power in low- and high-frequency EEG bands and impaired functional connectivity as common EEG abnormalities in autism spectrum disorders [20]. The organization of this article is as follows. In Section II, we briefly review some related work in autism diagnosis through EEG analysis. In Section III, the proposed system for preprocessing the EEG signal and extracting and classifying the EEG features of autism patients is presented. Then the obtained results are presented in Section IV. Section V discussed the obtained results. Finally, in the last section, the conclusion of this work is presented.

## II. RELATED WORK

As mentioned, due to the increasing prevalence of autism worldwide, the development of a new diagnostic approach for autism spectrum disorder is of great importance. Several studies have shown significant differences between various EEG features of individuals with autism and normal individuals [21-23]. These studies mainly focused on the spectral power of EEG frequency bands, connectivity and coherence between cortical areas, and hemispheric activity asymmetry [24, 25]. For instance, Jamal et al. obtained an accuracy, sensitivity and specificity of 94.7%, 85.7% and 100%, respectively, in autism diagnosis using connectivity features and a support vector machine (SVM) with the polynomial kernel [26]. Sheikhan et al. used spectral power features and K-nearest neighbours and reported a classification accuracy of 82.4% for autism diagnosis [27]. Fan et al.

proposed a similar framework for EEG classification of autism and non-autism subjects and reported an accuracy of 85% for this purpose [28].

On the other hand, several recent studies have investigated the nonlinear features of EEG for autism diagnosis. Bosl et al. reported good EEG classification results for autism diagnosis using multiscale entropy and SVM [29]. Ahmadlou et al. used the fractal dimension of EEGs and radial basis function neural network and achieved a classification accuracy of 90% for autism diagnosis [30]. Djemal et al. proposed an EEG-based computer-aided diagnosis of autism through entropy and wavelet-based features as well as an artificial neural network and achieved a classification accuracy of 99.71% [31]. Abdulhay et al. proposed a computer-aided autism diagnosis system through second-order plot area and empirical mode decomposition and achieved a classification accuracy of 94.4% for autism diagnosis [24].

Although several studies have been published in the field of computer-aided diagnosis of autism using EEG signals, very few studies have paid attention to the fact that in autism, we are dealing with a relatively wide range of patients with different behavioral and cognitive symptoms. Meanwhile, diagnosing the severity of autism is very important for clinicians in the treatment process. Therefore, in this study, we intend to classify the EEG signals of mild and severe autism patients using coherence indices and machine learning approaches (i.e., SVM and artificial neural networks).

### III. MATERIALS AND METHODS

#### A. Patients

Twelve patients with mild autism and twelve patients with severe autism were participated in the present research. All participants with autism were diagnosed by psychiatry experts according to the diagnostic criteria of DSM-5 [32]. Also, the severity of autism was determined through psychiatric interviews with specialists. The age range of the selected patients was 10-30 years. During the selection stage, patients with autism who had other neurological conditions, such as epilepsy or head trauma, were excluded. The patient enrollment was administered in a psychiatric clinic. The research project was done in accordance with the principles of the Declaration of Helsinki (1996) and the current Good Clinical Practice guidelines. The goal and an overview of the project were characterized by the participants and their parents during the initial contact. For those who agreed to participate, all the necessary information was provided prior to signing written informed consent. Information about the subjects was utilized anonymously and for the purpose of the study.

#### B. Data Acquisition and Cleaning

Previous studies have shown that resting with eyes open is the best EEG recording condition for EEG classification of autism from healthy subjects. Hence, in the current work, EEG recordings were performed while the patients were awake with their eyes open and sitting comfortably in an armchair without any stimuli. Depending on the subject's cooperation, EEG was recorded for 12-20 minutes for each patient in one session. Due to the difficulties of working with autism patients and recording EEG signals from these patients in the awake state,

the Emotiv Epoch headset device was utilized in this work. Since the Emotiv Epoch headset is a wireless EEG device, the signal recording was conducted in autistic patients more easily. This EEG device uses a Bluetooth module for wireless communication. The Emotiv Epoch headset and Software Development Kit include 14 electrodes (AF3, AF4, F7, F8, F3, F4, FC5, FC6, T7, T8, P7, P8, O1, O2 based on 10-20 international system) along with DRL/CMS references at P4/P3 locations. As depicted in Fig. 1, the EEG headset is wireless and has a large lithium-based battery for 12 hours. The sampling rate in this device is 128 Hz. The electrode impedance is reduced through the saline liquid and alcohol pads until the circles shown in Fig. 1 turn green. Emotive software was utilized to record EEGs and convert their format to MATLAB format.

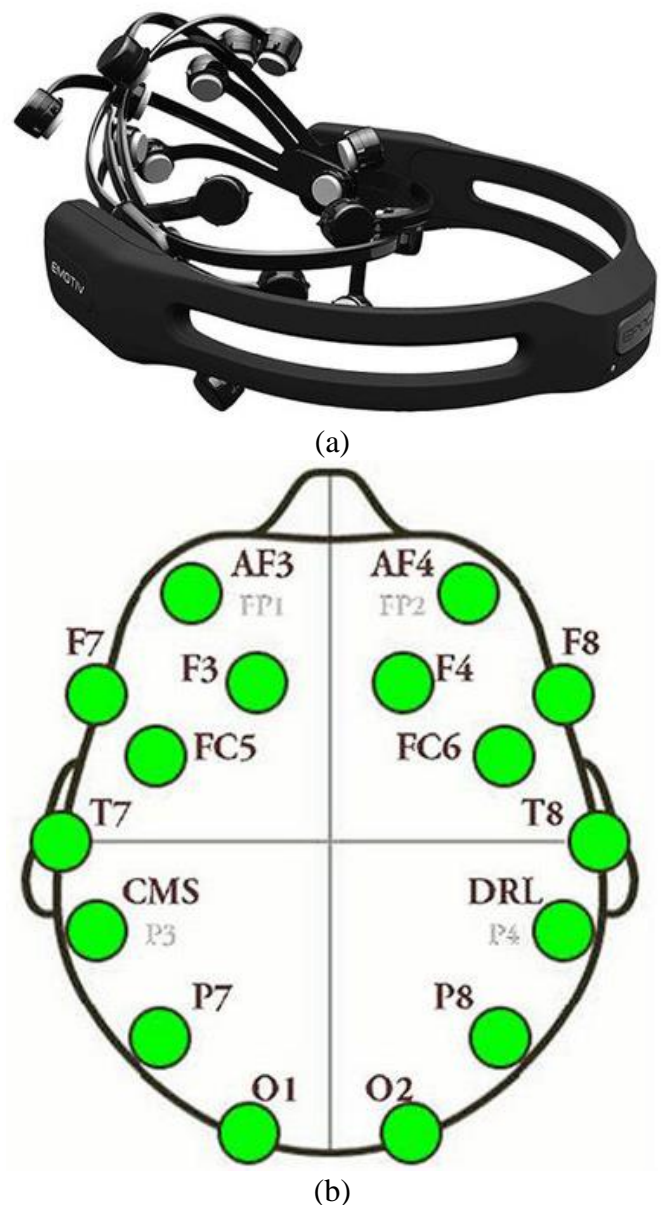


Fig. 1. Emotiv Epoch device and sensors organization. (a) headset and sensors, (b) 10-20 EEG international protocol.

Emotiv Epoch utilizes a 50 Hz notch filter to eliminate the main power line component. After EEG recording, in the signal preprocessing and conditioning stage, a band-pass Hanning window with a finite duration and frequency range of 1-45 Hz was applied to the signals via MATLAB software [33-35]. In addition, we performed electrode interpolation using adjacent channels for low-quality electrodes. EEGs were re-referenced to the common average and then were decomposed via independent component analysis (ICA). Components with motion and muscle artifacts were recognized by ICA and were then removed based on time courses and frequency scalp maps. The cleaned components were reconstructed, and a 50-second cleaned EEG signal was prepared for each patient.

### C. Coherence Features

EEG coherences measure the level of synchronization between two channels (i.e., two brain regions) in terms of EEGs recorded at different regions of the scalp. The coherence function gives information about the functional connectivity of the brain [36]. The value of the coherence function is in the range of zero to one, which shows the correlation of two signals in terms of frequency. If the value of the coherence function is zero, it means that the two channels are independent of each other, and the value of one indicates a high correlation between the two channels. The coherence function of two signals,  $i$  and  $j$ , at frequency  $f$ , is calculated as follows:

$$COH_{i,j}(f) = \frac{|S_{i,j}(f)|^2}{S_{i,i}(f) \cdot S_{j,j}(f)} \quad (1)$$

Where,  $S_{i,j}(f)$  is the cross-spectrum of the two signals, and  $S_{i,i}(f)$  and  $S_{j,j}(f)$  are auto-spectrum of the two signals. Fig. 2 shows the EEG coherence visualization method through matrix representation and node-link diagram.

In this work, we calculated short-range and long-range coherence values in the frequency range of 1-45 Hz, including short- and long-range intra- and inter-hemispheric coherence features. Totally, 89 coherence features were calculated.

### D. Feature Selection and Classification

To avoid the curse of dimensionality, statistical analysis was applied to the calculated coherence features to select features with significant differences between the two groups of patients with autism. Statistical analysis was also performed in MATLAB software. After performing the Shapiro-Wilk test to confirm the normality of the data, repeated measures analysis of variance (ANOVA) was used to handle the multiple comparison problems and the independent t-test was used as a post hoc analysis to compare the mean of the extracted features between the two groups [37-40].  $P < 0.05$  was considered as the threshold of significance.

### E. Multilayer Perceptron (MLP) Neural Network

The selected coherence features through a statistical analysis were applied to suitable classifiers to classify the feature set into mild and severe autism. SVM and artificial neural networks are common classifiers in biomedical applications. Neural networks have been utilized in classification and pattern recognition for different research projects because of their unique properties, such as self-organizing, adaptability, and robustness [41]. There are three layers in a feed-forward neural network like MLP: an input layer, a hidden layer, and an output layer. In the current research, the supervised learning network was utilized to classify mild and severe autism. The architecture of the network is depicted in Fig. 3.

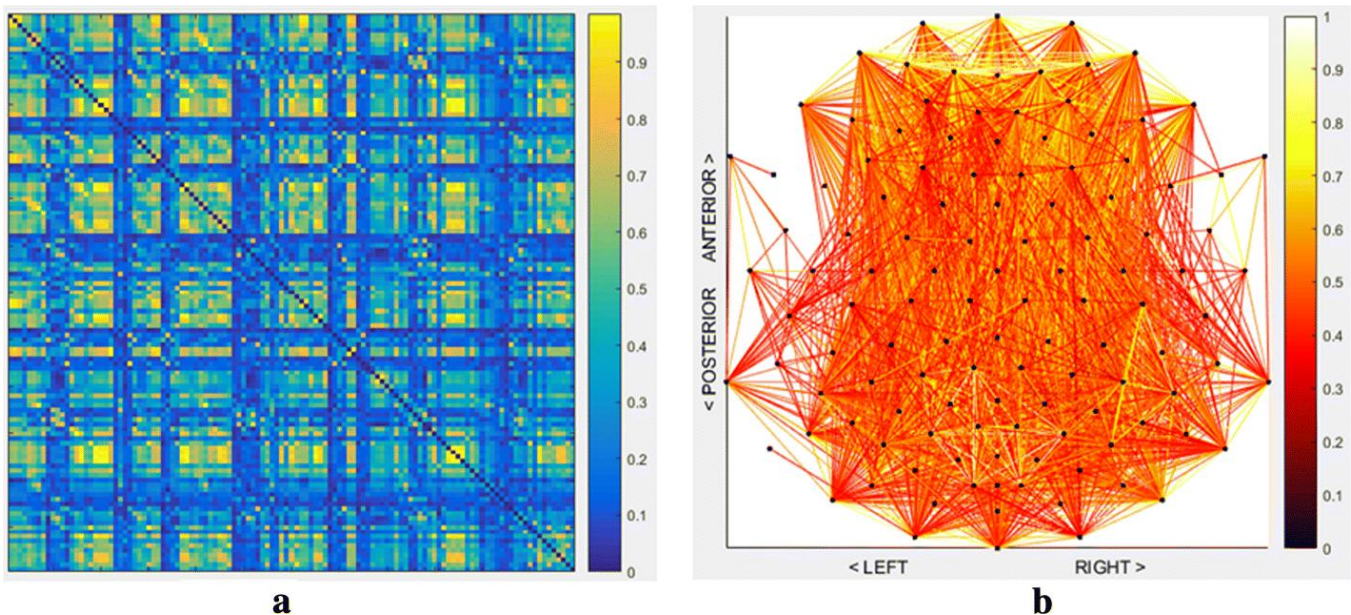


Fig. 2. EEG coherence visualization method. (a) matrix representation, (b) node-link diagram [43].

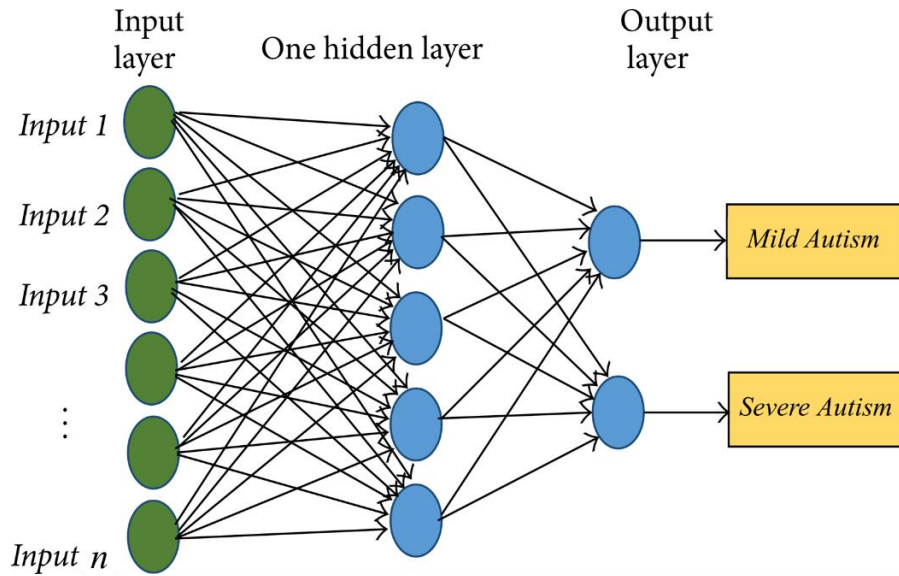


Fig. 3. Artificial neural network architecture.

The neural network was configured with  $n$  neurons (based on the number of selected features) in the input layer, five neurons in the hidden layer, hyperbolic tangent hidden transfer function, softmax output transfer function, and backward propagation training algorithm. The bias and weights for the network were initialized through the Nguyen-Widrow method. The neurons of each layer are connected to the next layer with a certain weight, which is defined as follows:

$$\Delta W_{ij} = \eta \delta_j(n) y_i(n) \quad (2)$$

The above equation is known as the delta law, through which weight correction is done from neuron  $i$  to neuron  $j$ .  $\eta$ ,  $\delta_j(n)$  and  $y_i(n)$  are the learning rate parameter, local gradient and input signal of neuron  $j$ , respectively. If  $j$  is a neuron in the hidden layer, then  $\delta_j(n)$  is obtained by:

$$\delta_j(n) = \varphi'_j(v_j(n)) \sum_k \delta_k(n) W_{kj}(n) \quad (3)$$

Where,  $k$  is a neuron in the output layer, and  $\varphi'_j(v_j(n))$  denotes the activation function to characterize the input-output relationships of the non-linearity to neuron  $j$ .

#### F. SVM with Radial Basis Function (RBF) Kernel

SVM has been widely utilized by researchers to solve various nonlinear problems and classification tasks with small data samples [42]. We used SVM in this study because this classifier minimizes the expected risk in the test data and considers a margin around the class boundaries, which leads to increased generalizability of the results. SVM uses a kernel function to transform the nonlinear classification problem into a linear one by increasing the dimensionality of the dataset. In this work, we used the RBF kernel. The RBF kernel is the most widely utilized kernel in SVMs. The RBF kernel has good performance in various classification problems. It is expressed as:

$$(x_i \cdot x_j) = \exp(-\gamma \|x_i - x_j\|^2) \cdot \gamma > 0 \quad (4)$$

Where,  $\gamma$  is the free parameter to scale the extent of influence, two samples have on each other.

#### IV. RESULTS AND EVALUATION

The coherence features were calculated from every 14 channels of EEG signals. Fig. 4 shows an example of recorded EEG signals of mild autistic and severe autistic patients for the right and left hemispheres after preprocessing. After feature extraction, the feature selection process was done through statistical analysis. Fig. 5 shows box plots for short-range and long-range intra-hemispheric coherence features with significant differences ( $P < 0.05$ ) between mild and severe autism groups. As shown, there were significant differences in both the left and right hemispheres. In addition, Fig. 6 shows box plots for short-range and long-range inter-hemispheric coherence features with significant differences ( $P < 0.05$ ) between mild and severe autism groups. These 10 coherence features with significant differences between the two groups were considered as selected features for the classification step. In the classification step, based on the results obtained from the statistical analysis, we considered four feature sets as input to the classifiers: all coherence features, all selected features, selected intra-hemispheric features, and selected inter-hemispheric features. It should be noted that the training and testing processes of the classifier were carried out with the four-fold cross-validation method.

Accuracy, sensitivity and specificity were calculated in this study to measure classification performance. These metrics were calculated based on the concepts of false negative (FN), false positive (FP), true negative (TN), and true positive (TP), which represent cases that were incorrectly or correctly identified as negative or positive cases.

$$\begin{aligned} \text{Accuracy} &= \frac{TP+TN}{N+P} \\ \text{Sensitivity} &= \frac{TP}{TP+FN} \\ \text{Spesificity} &= \frac{TN}{TN+FP} \end{aligned} \quad (5)$$

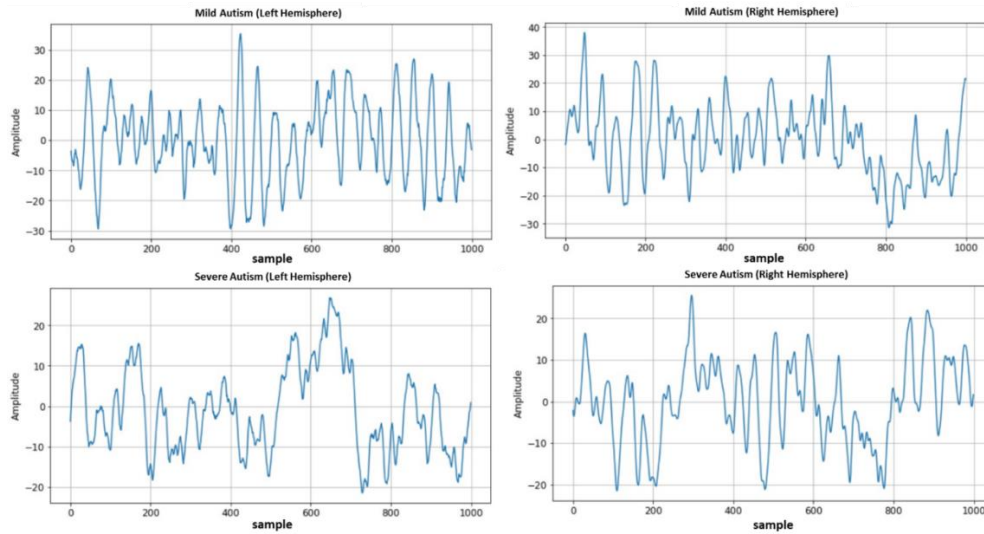


Fig. 4. Sample EEG signals of mild autistic (top) and severe autistic (bottom) patients for right and left hemispheres.

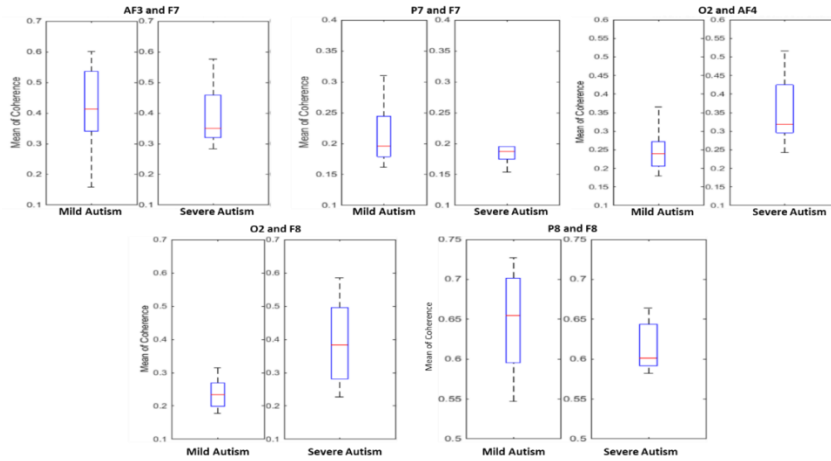


Fig. 5. Box plots for short- and long-range intra-hemispheric coherence features with significant differences ( $P < 0.05$ ) between mild and severe autism groups.

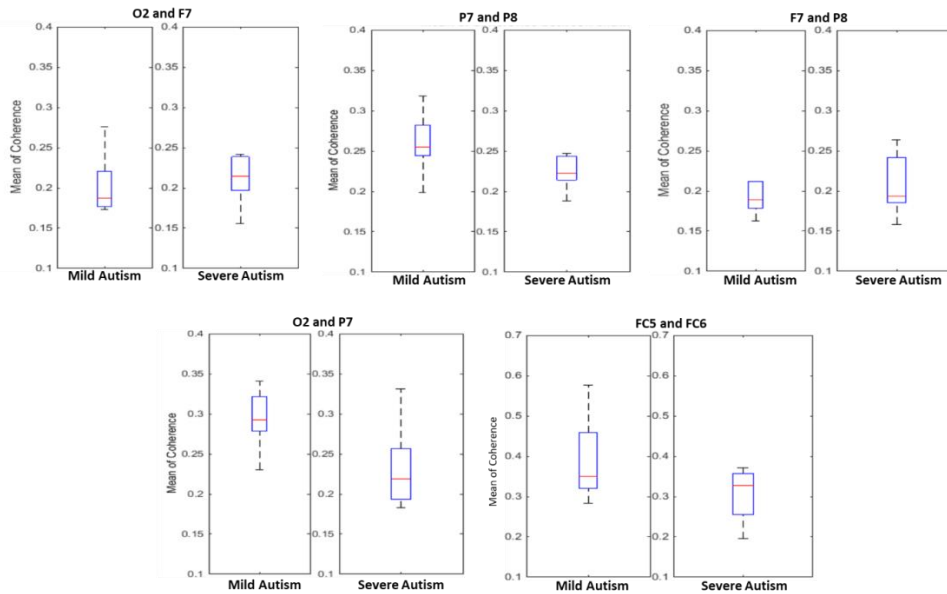


Fig. 6. Box plots for short- and long-range inter-hemispheric coherence features with significant differences ( $P < 0.05$ ) between mild and severe autism groups.

The performance of the MLP and SVM classifiers is evaluated based on the above indices, as depicted in Tables I and II.

Table I shows the performance of the MLP neural network in the classification of coherence features using the four mentioned feature sets. According to this table, the best classification performance was obtained by selected inter-hemispheric coherence features with accuracy, sensitivity and specificity of 96.82%, 97.82% and 96.92%, respectively. Furthermore, Table II shows the performance of the SVM classifier with RBF kernel in the classification of coherence features using the four mentioned feature sets. Again, the best classification performance was obtained by selected inter-

hemispheric coherence features with accuracy, sensitivity and specificity of 94.70%, 93.85% and 95.55%, respectively.

Fig. 7 compares the accuracy rates obtained by MLP and SVM for coherence features extracted from EEG signals for mild and severe autism classification. As shown, the MLP neural network has performed better in classifying coherence features and distinguishing mild autism from severe autism compared to the SVM classifier with the RBF kernel. However, it should be noted that the MLP neural network imposes a much higher computational cost than the SVM classifier. Table III shows the average time of classification operations (in terms of seconds) using MLP and SVM classifiers with and without feature selection.

TABLE I. PERFORMANCE OF MLP NEURAL NETWORK IN THE CLASSIFICATION OF COHERENCE FEATURES

Feature set	Accuracy (%)	Sensitivity (%)	Specificity (%)
All coherence features	88.94	85.90	89.36
All selected features	92.25	91.45	92.99
Selected intra-hemispheric features	93.68	93.70	92.51
Selected inter-hemispheric features	96.82	97.82	96.92

TABLE II. PERFORMANCE OF SVM CLASSIFIER WITH RBF KERNEL IN THE CLASSIFICATION OF COHERENCE FEATURES

Feature set	Accuracy (%)	Sensitivity (%)	Specificity (%)
All coherence features	86.39	84.10	87.91
All selected features	91.12	90.34	92.02
Selected intra-hemispheric features	91.97	89.63	92.69
Selected inter-hemispheric features	94.70	93.85	95.55

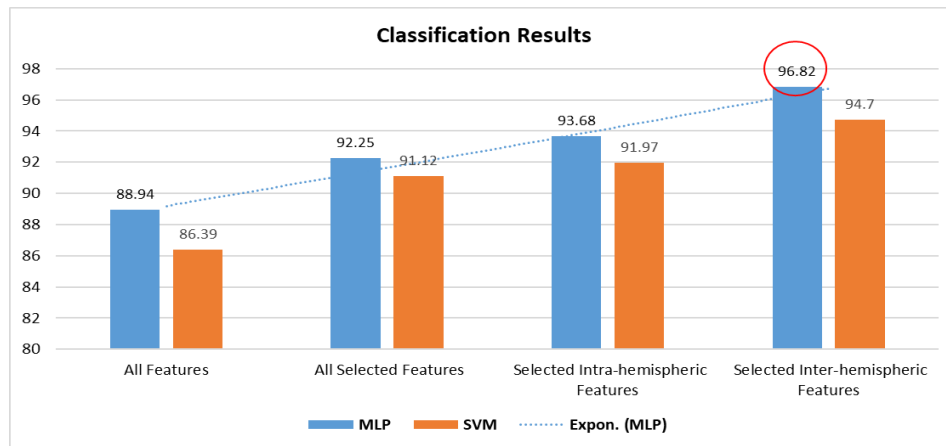


Fig. 7. Average accuracies obtained by MLP and SVM for coherence features extracted from EEG signals for mild and severe autism classification.

TABLE III. AVERAGE TIME OF CLASSIFICATION OPERATIONS USING MLP AND SVM CLASSIFIERS WITH AND WITHOUT FEATURE SELECTION

Operation	Time (s)
MLP classification without feature selection	40.82
MLP classification with feature selection	5.16
SVM classification without feature selection	0.62
MLP classification with feature selection	0.28



## V. DISCUSSION

In the present work, we attempted to investigate the potential of EEG coherence features in the classification of mild from severe autism and compare the ability of the MLP neural network and SVM classifier with RBF kernel to classify these features. Our findings showed that the MLP neural network has a better classification performance than the SVM classifier in the task of classifying the coherence features extracted from the EEG signals of two patient groups (96.82% versus 94.70%) at the cost of a more complicated computational process. The computation time of MLP classifier implementation was much longer than that of SVM. Very few similar studies have been published to date to classify mild from severe autism through brain signal analysis. For this reason, it is very challenging to compare the results of this research with other methods based on engineering algorithms. Cheong et al. applied wavelet transform to feature extraction from EEG signals and MLP classifier and reported an accuracy of 92.3% for classifying mild autistic patients from severe autistic patients [44]. In a three-class problem, Howell et al. [45] proposed a general linear model for feature extraction, a recursive feature elimination method for feature selection, and a random forest classifier to classify fMRI data into mild, moderate, and severe autism. They reported 72% accuracy on this three-class problem. Therefore, compared to the previous limited works in this field, our neural network-based system performs better. This can be due to the use of coherence features in the present work. Many electrophysiology and neuroimaging studies on autism spectrum disorders have shown that abnormality and impairment in brain connectivity are one of the most reported neuropathological mechanisms of autism [46-48]. Therefore, the coherence feature, which expresses the degree of coupling of different brain regions with each other, can be one of the strengths of our proposed system for classifying different severities of autism. In addition, it should be noted that most coherence impairments are located in the frontal region, consistent with previous neurophysiological works [49, 50].

Although SVM is a powerful classifier for two-class classification problems, MLP neural network was able to show better performance in the present work. However, the high computational cost of neural networks can limit their practical application in clinical systems [51-54]. Therefore, it is recommended to optimize the MLP neural network in order to reduce the calculation cost and improve the classification performance in future research. In addition, the investigation of other SVM kernels (such as linear, polynomial or sigmoid kernels) can also be done in future research. Furthermore, our results showed that feature selection through statistical analysis is a suitable approach to optimize the two-class classification problem of autism spectrum disorders. In fact, the feature selection approach adopted in this study led to improved classification performance and a significant reduction in computational cost. Therefore, future studies should do more research on the feature selection stage extracted from the EEG signals.

The strength of our study is to propose a simple semi-automatic system based on coherence features and a neural network for classifying mild autistic patients from severe

autistic patients from EEG signals. However, limitations such as the small sample size could reduce the generalizability of the results of the current research. Furthermore, in this work, the resting state EEG was analyzed, while previous neurophysiological studies have demonstrated that patients with autism spectrum disorders exhibit important neuropathological mechanisms in different states of arousal. Therefore, different EEG recording protocols should be considered in future studies.

## VI. CONCLUSION

In this paper, the classification of mild and severe autism from EEG signals was investigated by coherence features with MLP neural network and SVM classifier with RBF kernel. The effectiveness of these features was investigated via statistical analysis, and it was seen that coherence features with significant differences between the two groups have more discrimination in diagnosing autistic patients with different severities. In addition, the effectiveness of MLP and SVM was compared, and the MLP neural network yielded the maximum classification accuracy, sensitivity, and specificity. Considering that our simple system gives promising results in diagnosing autistic patients with mild and severe severities from EEG, there is scope for further work with a larger sample size and different ages and genders.

## REFERENCES

- [1] H. Dadgar, J. A. Rad, Z. Soleymani, A. Khorammi, J. McCleery, and S. Maroufizadeh, "The relationship between motor, imitation, and early social communication skills in children with autism," *Iranian journal of psychiatry*, vol. 12, no. 4, p. 236, 2017.
- [2] M. R. Mohammadi, H. Zarafshan, and S. Ghasempour, "Broader autism phenotype in Iranian parents of children with autism spectrum disorders vs. normal children," *Iranian journal of psychiatry*, vol. 7, no. 4, p. 157, 2012.
- [3] M. Al-Diabat, "Fuzzy data mining for autism classification of children," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 7, 2018.
- [4] N. Ghahari, F. Yousefian, S. Behzadi, and A. Jalilzadeh, "Rural-Urban Differences in Age at Autism Diagnosis: A Multiple Model Analysis," *Iranian Journal of Psychiatry*, vol. 17, no. 3, pp. 294-303, 2022.
- [5] M. R. Mohammadi et al., "Prevalence and correlates of psychiatric disorders in a national survey of Iranian children and adolescents," *Iranian journal of psychiatry*, vol. 14, no. 1, p. 1, 2019.
- [6] M. R. Mohammadi et al., "Prevalence of autism and its comorbidities and the relationship with maternal psychopathology: a national population-based study," *Arch Iran Med*, vol. 22, no. 10, 2019.
- [7] A. Khaleghi et al., "Epidemiology of psychiatric disorders in children and adolescents; in Tehran, 2017," *Asian journal of psychiatry*, vol. 37, pp. 146-153, 2018.
- [8] M. R. Alteneiji, L. M. Alqaydi, and M. U. Tariq, "Autism spectrum disorder diagnosis using optimal machine learning methods," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 9, 2020.
- [9] H. Zarafshan, M. R. Mohammadi, F. Abolhassani, S. A. Motevalian, and V. Sharifi, "Developing a comprehensive evidence-based service package for toddlers with autism in a low resource setting: Early detection, early intervention, and care coordination," *Iranian Journal of Psychiatry*, vol. 14, no. 2, p. 120, 2019.
- [10] C. Ecker and D. Murphy, "Neuroimaging in autism—from basic science to translational research," *Nature Reviews Neurology*, vol. 10, no. 2, pp. 82-91, 2014.
- [11] R. Chen, Y. Jiao, and E. H. Herskovits, "Structural MRI in autism spectrum disorder," *Pediatric research*, vol. 69, no. 8, pp. 63-68, 2011.

- [12] S. Jebapriya, D. Shubin, J. W. Kathrine, and N. Sundar, "Support vector machine for classification of autism spectrum disorder based on abnormal structure of corpus callosum," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 9, 2019.
- [13] M. J. Walsh, G. L. Wallace, S. M. Gallegos, and B. B. Braden, "Brain-based sex differences in autism spectrum disorder across the lifespan: A systematic review of structural MRI, fMRI, and DTI findings," *NeuroImage: Clinical*, vol. 31, p. 102719, 2021.
- [14] A. Khorrami, M. Tehrani-Doost, and H. Esteky, "Comparison between face and object processing in youths with autism spectrum disorder: an event related potentials study," *Iranian Journal of Psychiatry*, vol. 8, no. 4, p. 179, 2013.
- [15] A. Khaleghi, M. R. Mohammadi, K. Shahi, and A. M. Nasrabadi, "Computational neuroscience approach to psychiatry: A review on theory-driven approaches," *Clinical Psychopharmacology and Neuroscience*, vol. 20, no. 1, p. 26, 2022.
- [16] S. Alhagry, A. A. Fahmy, and R. A. El-Khoribi, "Emotion recognition based on EEG using LSTM recurrent neural network," *International Journal of Advanced Computer Science and Applications*, vol. 8, no. 10, 2017.
- [17] A. Al-Nafjan, M. Hosny, A. Al-Wabil, and Y. Al-Ohali, "Classification of human emotions from electroencephalogram (EEG) signal using deep neural network," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 9, pp. 419-425, 2017.
- [18] M. A. Elshahed, "Towards using Single EEG Channel for Human Identity Verification," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 5, 2021.
- [19] H. Hadoush, M. Alafeef, and E. Abdulhay, "Brain complexity in children with mild and severe autism spectrum disorders: analysis of multiscale entropy in EEG," *Brain topography*, vol. 32, pp. 914-921, 2019.
- [20] J. Wang, J. Barstein, L. E. Ethridge, M. W. Mosconi, Y. Takarae, and J. A. Sweeney, "Resting state EEG abnormalities in autism spectrum disorders," *Journal of neurodevelopmental disorders*, vol. 5, pp. 1-14, 2013.
- [21] L. M. Oberman, E. M. Hubbard, J. P. McCleery, E. L. Altschuler, V. S. Ramachandran, and J. A. Pineda, "EEG evidence for mirror neuron dysfunction in autism spectrum disorders," *Cognitive brain research*, vol. 24, no. 2, pp. 190-198, 2005.
- [22] A. Yasuhara, "Correlation between EEG abnormalities and symptoms of autism spectrum disorder (ASD)," *Brain and Development*, vol. 32, no. 10, pp. 791-798, 2010.
- [23] T. A. Stroganova et al., "Abnormal EEG lateralization in boys with autism," *Clinical Neurophysiology*, vol. 118, no. 8, pp. 1842-1854, 2007.
- [24] E. Abdulhay et al., "Computer-aided autism diagnosis via second-order difference plot area applied to EEG empirical mode decomposition," *Neural Computing and Applications*, vol. 32, pp. 10947-10956, 2020.
- [25] N. Pop-Jordanova, T. Zorcec, A. Demerdzieva, and Z. Gucev, "QEEG characteristics and spectrum weighted frequency for children diagnosed as autistic spectrum disorder," *Nonlinear Biomedical Physics*, vol. 4, pp. 1-7, 2010.
- [26] W. Jamal, S. Das, I.-A. Oprescu, K. Maharatna, F. Apicella, and F. Sicca, "Classification of autism spectrum disorder using supervised learning of brain connectivity measures extracted from synchronostates," *Journal of neural engineering*, vol. 11, no. 4, p. 046019, 2014.
- [27] H. Behnam, A. Sheikhan, M. Noroozian, and M. R. Mohammadi, "Abnormalities of Quantitative Electroencephalography in Children with Asperger Disorder Using Spectrogram and Coherence Values," *Iranian Journal of Psychiatry*, vol. 3, no. 2, pp. 64-70, 2008.
- [28] J. Fan et al., "A Step towards EEG-based brain computer interface for autism intervention," in *2015 37th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, 2015: IEEE, pp. 3767-3770.
- [29] W. Bosl, A. Tierney, H. Tager-Flusberg, and C. Nelson, "EEG complexity as a biomarker for autism spectrum disorder risk," *BMC medicine*, vol. 9, no. 1, pp. 1-16, 2011.
- [30] M. Ahmadi, H. Adeli, and A. Adeli, "Fractality and a wavelet-chaos-neural network methodology for EEG-based diagnosis of autistic spectrum disorder," *Journal of Clinical Neurophysiology*, vol. 27, no. 5, pp. 328-333, 2010.
- [31] R. Djemal, K. AlSharabi, S. Ibrahim, and A. Alsuwailem, "EEG-based computer aided diagnosis of autism spectrum disorder using wavelet, entropy, and ANN," *BioMed research international*, vol. 2017, 2017.
- [32] D. American Psychiatric Association and A. P. Association, *Diagnostic and statistical manual of mental disorders: DSM-5 (no. 5)*. American psychiatric association Washington, DC, 2013.
- [33] A. Khaleghi, A. Sheikhan, M. R. Mohammadi, and A. M. Nasrabadi, "Evaluation of cerebral cortex function in clients with bipolar mood disorder I (BMD I) compared with BMD II using QEEG analysis," *Iranian Journal of Psychiatry*, vol. 10, no. 2, p. 93, 2015.
- [34] M. Moeini, A. Khaleghi, N. Amiri, and Z. Niknam, "Quantitative electroencephalogram (QEEG) spectrum analysis of patients with schizoaffective disorder compared to normal subjects," *Iranian Journal of Psychiatry*, vol. 9, no. 4, p. 216, 2014.
- [35] M. Moeini, A. Khaleghi, and M. R. Mohammadi, "Characteristics of alpha band frequency in adolescents with bipolar II disorder: a resting-state QEEG study," *Iranian journal of psychiatry*, vol. 10, no. 1, p. 8, 2015.
- [36] A. Khaleghi, P. M. Birgani, M. F. Fooladi, and M. R. Mohammadi, "Applicable features of electroencephalogram for ADHD diagnosis," *Research on Biomedical Engineering*, vol. 36, pp. 1-11, 2020.
- [37] W. Xiao, G. Manyi, and A. Khaleghi, "Deficits in auditory and visual steady-state responses in adolescents with bipolar disorder," *Journal of Psychiatric Research*, vol. 151, pp. 368-376, 2022.
- [38] H. Zarafshan, A. Khaleghi, M. R. Mohammadi, M. Moeini, and N. Malmir, "Electroencephalogram complexity analysis in children with attention-deficit/hyperactivity disorder during a visual cognitive task," *Journal of clinical and experimental neuropsychology*, vol. 38, no. 3, pp. 361-369, 2016.
- [39] A. Khaleghi, M. R. Mohammadi, M. Moeini, H. Zarafshan, and M. Fadaei Fooladi, "Abnormalities of alpha activity in frontocentral region of the brain as a biomarker to diagnose adolescents with bipolar disorder," *Clinical EEG and neuroscience*, vol. 50, no. 5, pp. 311-318, 2019.
- [40] A. Khaleghi, H. Zarafshan, and M. R. Mohammadi, "Visual and auditory steady-state responses in attention-deficit/hyperactivity disorder," *European archives of psychiatry and clinical neuroscience*, vol. 269, pp. 645-655, 2019.
- [41] M. R. Mohammadi, A. Khaleghi, A. M. Nasrabadi, S. Rafieivand, M. Begol, and H. Zarafshan, "EEG classification of ADHD and normal children using non-linear features and neural network," *Biomedical Engineering Letters*, vol. 6, pp. 66-73, 2016.
- [42] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, pp. 273-297, 1995.
- [43] C. Ji, N. M. Maurits, and J. B. Roerdink, "Data-driven visualization of multichannel EEG coherence networks based on community structure analysis," *Applied Network Science*, vol. 3, no. 1, pp. 1-24, 2018.
- [44] L. C. Cheong, R. Sudirman, and S. S. Hussin, "Feature extraction of EEG signal using wavelet transform for autism classification," *ARPJ Journal of Engineering and Applied Sciences*, vol. 10, no. 19, pp. 8533-8540, 2015.
- [45] R. Haweel et al., "A novel framework for grading autism severity using task-based fmri," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, 2020: IEEE, pp. 1404-1407.
- [46] A. Sheikhan, H. Behnam, M. Mohammadi, and M. NOUROUZIAN, "Evaluation of quantitative electroencephalography in children with autistic disorders in various conditions based on spectrogram," 2008.
- [47] I. Mohammad-Rezazadeh, J. Frohlich, S. K. Loo, and S. S. Jeste, "Brain connectivity in autism spectrum disorder," *Current opinion in neurology*, vol. 29, no. 2, p. 137, 2016.
- [48] S. Wass, "Distortions and disconnections: disrupted brain connectivity in autism," *Brain and cognition*, vol. 75, no. 1, pp. 18-28, 2011.
- [49] M. A. Just, T. A. Keller, V. L. Malave, R. K. Kana, and S. Varma, "Autism as a neural systems disorder: a theory of frontal-posterior underconnectivity," *Neuroscience & Biobehavioral Reviews*, vol. 36, no. 4, pp. 1292-1313, 2012.

- [50] E. L. Hill, "Executive dysfunction in autism," *Trends in cognitive sciences*, vol. 8, no. 1, pp. 26-32, 2004.
- [51] A. Afzali, A. Khaleghi, B. Hatef, R. Akbari Movahed, and G. Pirzad Jahromi, "Automated major depressive disorder diagnosis using a dual-input deep learning model and image generation from EEG signals," *Waves in Random and Complex Media*, pp. 1-16, 2023.
- [52] A. Khaleghi, M. R. Mohammadi, G. P. Jahromi, and H. Zarafshan, "New ways to manage pandemics: Using technologies in the era of covid-19: A narrative review," *Iranian journal of psychiatry*, vol. 15, no. 3, p. 236, 2020.
- [53] A. Khaleghi, M. R. Mohammadi, K. Shahi, and A. Motie Nasrabadi, "A neuronal population model based on cellular automata to simulate the electrical waves of the brain," *Waves in Random and Complex Media*, pp. 1-20, 2021.
- [54] A. Khaleghi, M. R. Mohammadi, K. Shahi, and A. M. Nasrabadi, "Possible Neuropathological Mechanisms Underlying the Increased Complexity of Brain Electrical Activity in Schizophrenia: A Computational Study," *Iranian Journal of Psychiatry*, pp. 1-7, 2023.

# Enhancing Breast Cancer Diagnosis using a Modified Elman Neural Network with Optimized Algorithm Integration

Linkai Chen\*, CongZhe You, Honghui Fan, Hongjin Zhu

Jiangsu University of Technology, School of Computer and Engineering, Zhongwu Road No.1801, Changzhou, China, 2130011

**Abstract**—Breast cancer is a class of cancer that starts in the cells of the breast. It happens once the cells of the breast divide and amplify abnormally and uncontrollably. Other parts of the body, including lymph nodes, bones, lungs, and liver, can be affected by breast cancer. Early diagnosis and treatment are critical in helping to lessen the risk of death from breast cancer. Machine learning is a type of artificial intelligence that can be used to diagnose breast cancer. It uses algorithms to analyze data and assess patterns associated with breast cancer. Machine learning models can help improve diagnostic accuracy, reduce false-positive results, and improve the efficiency of diagnosis. Elman Neural Networks (ENNs) are machine learning algorithms that can be used to diagnose breast cancer. ENNs use medical data to detect patterns that are associated with the presence of cancer. The accuracy of ENNs in diagnosing breast cancer is still being researched, but they have the potential to help improve diagnostic accuracy and reduce false-positive results. In the existing study, a new modified version of ENN founded on a combination of an upgraded version of the imperialist competitive algorithm is proposed for this objective. Likewise, the results of the model compared with some other methods illustrated the proposed method's higher efficiency.

**Keywords**—Breast cancer model; Elman Neural Network; upgraded imperialist competitive algorithm

## I. INTRODUCTION

A class of cancer that begins in the cells of the breast is called breast cancer. It occurs once the cells in the breast start to expand abnormally and uncontrolled. It can affect both men and women, although women are more likely to be diagnosed with it than men [1]. The advantage of recognizing breast cancer in the initial steps is that it can significantly increase a person's chances of survival and decrease the risk of complications from surgery and other treatments [2]. Early treatment is also more likely to be successful, as tumors are generally smaller when detected in their earliest stages. Furthermore, early detection can reduce the risk of cancer spreading to other organs in the body.

The use of AI, machine learning, and computer-aided detection (CAD) technologies can help with early breast cancer diagnosis [3]. The CAD utilizes algorithms to pinpoint potentially harmful areas in mammography images and compare them with existing patterns [4]. Moreover, AI-based automatic identification of tumors is becoming more precise and is allowing doctors to detect smaller, possibly more aggressive cancers sooner. Additionally, AI is the wing of

computer science focused on developing machines able to execute duties that usually demand human wit. ML (Machine learning) is a type of AI which involves algorithms and models that can “learn” from data to make decisions on their own [5].

An ANN (artificial neural network) is a type of ML algorithm inspired by how neurons work in the human brain to process information. It uses a large number of connected “neurons” to form an interconnected web of nodes [6]. Each node can take in data, process it, and pass it along to another node, eventually forming a prediction or classification.

Vijayakumar et al. suggested a technique based on DNN (deep Feed-forward Neural Network) with four Activation Functions (AFs) for the category of breast cancer [7]. Its purpose was twofold: increase understanding of how different AFs can be used in different layers of a DNN and develop a predictive model with improved accuracy. The model performance was evaluated using a 10-fold Cross Validation (CV) method and various metrics. Results showed that the proposed solution performed better than other AFs-based DNNs, making it a viable expert-level system for breast cancer dataset classification [6]. Also, the method provided better results than the other validated techniques.

Al-Haija et al. introduced a precise and comprehensive computational model for diagnosing breast cancer with the help of ResNet-50 CNN [8]. The given framework used the ResNet-50 CNN as a pre-trained neural network on ImageNet for the purpose of transferring learning to classify the BreakHis dataset into benign or malignant. The evaluation results demonstrated that their method attained outstanding classification accuracy, exceeding the performance of other models trained on a similar dataset.

Ahila et al. presented a bio-inspired algorithm to optimize variables of a neural network for computer-aided diagnosis of breast ultrasound images [9]. The preprocessing of the images involves sigmoid filtering, despeckling, and anisotropic diffusion, followed by extraction of the location of concern from which weave and morphological features are tallied. The classification task is achieved using a wavelet neural network tuned with grey wolf optimization, culminating in evaluating the model's performance against 346 images via the receiver operating and confusion matrix characteristic. Analysis shows this method produces higher accuracy than existing methods, thus proving its application in accurate tumor detection and classification. Among different kinds of artificial neural

networks, Elman Neural Network (ENN) is an artificial Recurrent Neural Network (RNN) type of Artificial Intelligence (AI) that can be used to analyze and identify patterns in data. It has a recurrent connection between its input and output layers, allowing it to remember the previous output when computing its current output. This memory enables the algorithm to learn from past experiences and apply them to future decisions. In this study, a modified form of ENN (Elman Neural Network) founded on a modified metaheuristic has been used in order to Model Breast Cancer. The method is a variation of the Elman Neural Network specifically adapted for breast cancer modelling. The modifications are designed to enable the algorithm to analyze hundreds of different features of the breast tissue in order to identify cancerous cells accurately. In this process, the modified Elman Neural Network can model the data.

In summary, this paper delves into the pivotal role of AI, machine learning, and neural networks in revolutionizing the early detection and accurate classification of breast cancer. Through an exploration of various models, including Elman Neural Networks, ResNet-50 CNNs, and optimized algorithms, we aim to illuminate the strides being made towards enhancing diagnostic accuracy and ultimately improving patient outcomes. The subsequent sections delve deeper into these methodologies, their applications, and the promising results they yield. By harnessing the power of technology and innovative approaches, we aim to contribute to the ongoing efforts in combating breast cancer and transforming the landscape of medical diagnosis.

## II. ELMAN NEURAL NETWORKS

ENNs are a Recurrent Neural Network developed by Ronald J. Elman in the late 1980s [10]. Unlike traditional feed-forward neural networks, Elman Neural Networks can retain and use information from previous inputs, allowing them to understand context and patterns over time better. This makes them particularly well-suited for processing time series data [11]. They have been used in various applications, such as predicting stock market prices and recognizing speech patterns.

The core components of the network are analogous to those in a feed-forward neural network [12]; the junctions between the input layer ( $W_h^i$ ), the hidden layer ( $W_h^h$ ), and the output layer ( $W_h^o$ ) are similar to what's found in a multi-layer neural network. As seen in Fig. 1, this is an overall representation of an Elman Neural Network.

Furthermore, Elman Neural Networks includes an extra layer known as the context layer ( $W_h^c$ ), which takes input from the hidden layer's outputs in order to store the values of the hidden layer from the prior step [11]. This is visible in Fig. 1.

It is assumed that the dimensions of both the output and input layers are  $n$ , such that:

$$x^1(t) = [x_1^1(t), x_2^1(t), \dots, x_n^1(t)]^T \quad (1)$$

$$y(t) = [y_1(t), y_2(t), \dots, y_n(t)]^T \quad (2)$$

While the context layer dimension is  $m$ .

The  $l^{th}$  input layer and the  $k^{th}$  hidden layers are undeniably essential for optimal functionality.

$$u_i(l) = e_i(l), \quad (3)$$

$$v_k(l) = \sum_{j=1}^N \omega_{kj}^1(l)x_j^c(l) + \sum_{i=1}^n \omega_{ki}^2(l)u_i(l) \quad (4)$$

where,  $i = 1, 2, \dots, n$  and  $k = 1, 2, \dots, N$ .

The  $\omega_{kj}^l(l)$  describes the hidden layers' weights from the  $o^{th}$  node and  $x_j^c(l)$  signals converted from the  $k^{th}$  the node of the context layer powerfully demonstrates that there should be. Therefore, the weight of the hidden layer  $k$  for the input layer  $i$  can be determined by  $\omega_{kj}^2(l)$ .

Therefore, the following approach can determine the last output of the hidden layer presented to the context layer.

$$W_k(l) = f_o(\bar{v}_k(l)) \quad (5)$$

where,

$$\bar{v}_k(l) = \frac{1}{\max(v_k(l))} \times v_k(l) \quad (6)$$

where,  $\bar{v}_k(l)$  specifies the hidden layer's normal value.

As a result, the context layer produces its output in the following way.

$$C_k(l) = \beta \times C_k(l-1) + W_k(l-1), \quad (7)$$

Such that  $W_k|_{k=1,2,\dots,N}$  defines the self-connected feedback between  $[0, 1]$ .

As a result, the Elman Neural Network's output layer has been gained as follows:

$$y_o(l) = \sum_{k=1}^N \omega_{ok}^3(l)W_k(l), \quad (8)$$

Such that,  $\omega_{zk}^3(l)$  specifies the connection weight from the  $k^{th}$  layer into the  $z^{th}$  layer, where,  $z = 1, 2, \dots, n$ .

Utilizing Ren et al.'s method [13], the structure of the ENN has been enhanced. This approach is executed according to the pseudocode provided in Table I.

Such that  $c$  is a constant,  $t$  specifies the recent iteration,  $\mu$  signifies the learning rate.

To maximize the effectiveness of the upgraded Elman Neural Network, it is essential to incorporate it with a metaheuristic algorithm. In this study, an upgraded version of the imperialist competitive algorithm has been utilized for this purpose.

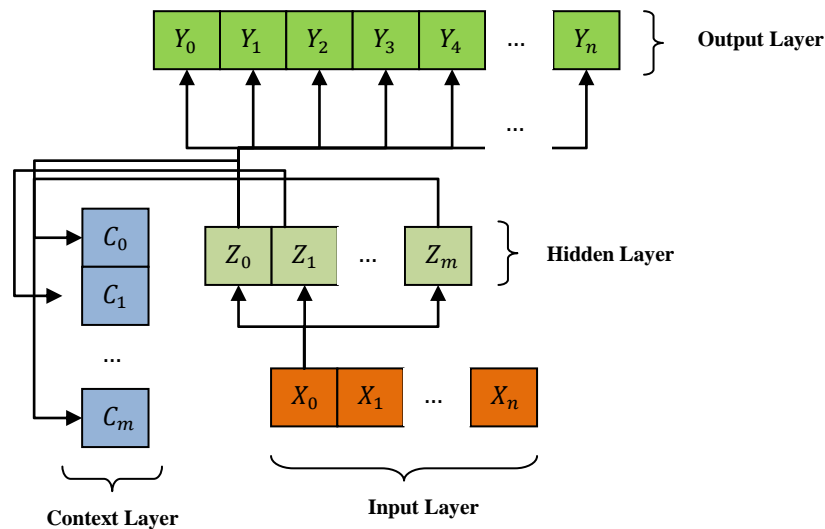


Fig. 1. Configuration of the Elman Neural Networks.

TABLE I. PSEUDOCODE OF THE IMPROVED ENN ON THE BASIS OF REN ET AL.'S METHOD [13]

Learning rate value initialization $\mu = \varepsilon$
<pre> In the lth iteration: if <math>l \leq 2</math> : <math>\mu = defaultvalue</math>; end if <math>t \geq 3 \&amp;\&amp; e(l) &lt; 1.02e(1 - 1)</math>: <math>\mu = \frac{c(1 + \frac{1}{l})^l}{exp(1)}</math> else <math>\mu = defaultvalue</math> end end Improve weights; compute <math>e(t)</math>; If stop criteria have been earned: stop; else carry on                     </pre>

### III. UPGADED IMPERIALIST COMPETITIVE ALGORITHM

#### A. The Conception of the Imperialist Competition Algorithm

As a population-based optimization approach, the Imperialist competition algorithm is introduced to solve optimization problems that take inspiration from human societal and cultural processes [14]. It begins with an initial population of random values and progresses toward an optimal solution by exploring the problem space, much like other population-based optimization methods.

This algorithm uses principles of assimilation, imperialistic competition, and revolution to solve complex optimization issues [15]. It takes the form of countries and gradually optimizes the solutions through an iterative process until it finds the best solution. As with other heuristic-based algorithms, it initiates with a haphazard initial population named "countries". A few of the highest-performing elements, akin to the GA elites, are then picked as "imperialists," while

the remainder is considered colonies [16]. These colonies were drawn to the imperialist based on their strength via a specific process. Any empire's power is determined by its imperialist country (as the central core) and its colonies.

The initial stage of the algorithm includes setting up the conditions for displaying the solution, generating the first group of entities, and forming the original colonies. The response will be shown as a vector with  $n$  components, where  $n$  is the quantity of orders and  $g$  is the number of pieces per order. Any piece must be manufactured  $n$  times, meaning it is repeated in each part number of the string [17]. Assembling begins by taking all parts that were created in the preliminary phase of production. The starting population of the ICA (the quantity of countries) is randomly generated. The algorithm stops when the set processing time elapses.

In order to ensure that a solution is always possible in the permutation display method for this problem, it is essential to observe the entire permutation and keep any operators from disrupting its completion [18]. To figure out the costs of each

country, their cost function is measured. From the number of empires, the participants in the candidates with the less cost function value are chosen as the remaining colonizers. The "Roulette wheel selection" technique is employed to split the colonies among the colonizers. This entails taking the cost of all colonists and computing their normalized cost with Eq. (9).

$$C_n = \exp(-c_n / \max c_i) \quad (9)$$

where,  $c_i$  describes the  $n^{th}$  imperialist cost value, and  $C_n$  signifies the  $n^{th}$  colonizer and  $n$  specify the normalized cost  $i$  of the maximum cost among all colonizers.  $n_E$  describes the number of empires. With  $\max c_i$  having a normalized cost, the normalized relative power of any colonizer is computed according to Eq. (10) and is the basis for dividing the colonized countries among the colonizers.

$$p_n = \left| \frac{c_n}{\sum_{i=1}^{n_E} c_i} \right| \quad (10)$$

And

$$\sum_{i=1}^{n_E} P_i = 1 \quad (11)$$

The roulette cycle technique is one of the most well-known selection approaches. Initially, the selection chance figures are organized nearby one another, and afterwards, a haphazard number in the scope of zero to one is made. This scope is chosen on the grounds that the aggregate of the determinations in likelihood will dependably be equivalent to one. The random number is compared to the roulette wheel interval to determine the colonizer that corresponds to it. Since the probability of each colonizer taking up a section of the roulette wheel, higher quality colonizers (with lower cost functions) are more likely to be picked. After assigning all countries to their respective colonizers, the process moves into a loop with the colonial competition algorithm, continuing until a set stopping condition is reached.

In every empire, the colonizing country tries to increase the number of its colonies in order to increase its influence; therefore, in each empire, the colonized countries move towards the respective colonizer. In the presented algorithm, the attraction policy is defined as follows: to change the colony based on the emperor, two points in the colony are selected and before and after these two points are removed; then the removed components are added to the solution based on their relative order in the colonizer.

The revolution brings about an abrupt alteration to a nation's social and political qualities. In the imperialist competition algorithm, revolution is simulated by moving an imperial country to a separate random position. From a computational standpoint, the revolution halts the evolutionary moves from getting stuck in a nearby optimum snare, potentially enhancing the spot of a nation and carrying it towards areas with a more desirable situation. In the presented algorithm, each colony within a respective empire may revolt with a designated likelihood.

In this Algorithm, 2 positions are randomly chosen and swapped. As a result of the migration to a colonizing nation and the implementation of policies of revolution, some of these countries might potentially move up in comparison to the

colonizer. This means that the colonizer and the colonized will exchange places with one another, and the algorithm will continue with the newly appointed ruling nation. This country can then begin to enforce assimilation policies on its colonies.

The might of the colonizing country as well as a ratio of the entire might of its colonies, is the entire power of an empire; Hence, the whole cost of an empire is calculated through Eq. (12):

$$TC_n = c_n + \zeta[\text{mean}(c_n^i)] \quad (12)$$

where,  $TC_n$  describes the whole cost of the  $n^{th}$  empire and  $\text{mean}(c_n^i)$  signifies the mean amount of the cost of the colonizer  $n$ ,  $\zeta$  is a positive integer. Any empire that cannot improve its power and lose its competitive power during the colonizer's competition will be eliminated. For modelling this fact, it is supposed that the empire being deleted is the feeble-existent empire. Therefore, in each iteration of the algorithm, one or a number of the weakest colonies are removed from the weakest empire, and a rivalry is created among every empire to take over these colonies. Noteworthy colonies will not necessarily be taken over by the strongest emperor. Rather, stronger empires are more likely to take over. For simulating the rivalry between empires to take over this colony, first of all, its normalized total cost is determined from the entire cost of the empire via Eq. (13):

$$NTC_n = \exp(-TC_n / \max TC_i) \quad (13)$$

$NTC_n$  describes the normalized total cost of the  $n^{th}$  empire,  $TC_n$  is the  $n^{th}$  empire, and  $\max TC_i$  is the total cost of the weakest empire.

With the normalized total cost, the probability of taking over the competing colony by each empire is calculated through Eq. (14):

$$P_{emp_n} = \frac{NTC_n}{\sum_{i=1}^{n_E} NTC_i} \quad (14)$$

With the possibility of taking over any empire, the said colony belongs to the empire that wins the Roulette wheel. In the proposed algorithm, when an empire having no colonies is added as a colony to another empire. Selection Empire is made through the roulette cycle.

This research introduces a modified version of the imperialist competitive algorithm (ICA) called the upgraded imperialist competition (UIC) algorithm that hopes to address the issues of being prone to local minima and slow convergence speeds. Despite its successes, the original algorithm still poses a challenge.

### B. Upgraded Imperialist Competition Algorithm

To avoid becoming stuck in a local optimum, population-based evolutionary algorithms can apply the concept of revolution (colonial competition) or mutation (genetic algorithm). This research specifically investigates the use of the Quad Countries Algorithm and its "rejection policy" in order to help populations escape from such traps.

The algorithm of the Quad Countries Algorithm is based on the ICA, and in addition to colonial and colonizing countries, two new types of countries with the names of independent

countries and countries seeking independence have been added to its set of countries [19]. Independent countries search the solution space randomly, and countries seeking independence distance themselves from colonizing countries according to a targeted policy called exclusionary policy.

In the basic Imperialist competitive algorithm, when the revolution operator is applied, firstly, the position of a number of colonial countries changes randomly, and secondly, it changes in the form of divergence (moving away from the colonizer).

While in the ICA, inspired by the policy of repelling the algorithm of the Quad Countries Algorithm, at the time of the revolution, firstly, the position of some colonial countries changed in a purposeful way, and secondly, it changes in the form of convergence (getting closer to the colonizer).

The purposeful calculation of the new position of the colonial countries when the revolution is applied is that the total distance between the colonizer and the target country is first calculated in all dimensions.

Then by multiplying the current position of the desired colony country by the resulting vector (Center), the desired country moves to a new position. This causes when the revolution operator is applied, the position of a percentage of the population elements is purposefully moved in the direction of the best element (the colonizer) to change.

$$Center = \sum(Col_i - Imp_j) \quad (15)$$

Eq. (15) shows how to calculate the Center vector. In this regard, first, the sum of the distances of all colony countries ( $Col$ ) with the colonizing empires ( $Imp$ ) is calculated. Then the product of this vector with a coefficient of 0.1 in the vector of the current position of the colony gives the new position of the colony country.

$$Col_i^{new} = Col_i^{old} \times (0.01 \times Center) \quad (16)$$

In the basic colonial competition algorithm, when a revolution occurs, the colonies move randomly in space and away from the colonizer, the best element of the empire, hoping to gain a better position in the problem space. While the colonies in the improved colonial competition algorithm avoid random movements while analyzing the movement of the best element of the group, they follow its movement and in order to reach a better position, not only do they not move away from the colonizer, but also according to their distance from the emperor. The leader's movement is inspired, and they step in that direction. This method makes all the colonies participating in the revolution process guess the optimal global location in the same number of algorithm iterations and converge around the global optimal point.

#### IV. OPTIMAL MODIFIED ENN USING UIC ALGORITHM

In the given section, the procedure for refining the ENN has been outlined. To refine the suggested ENN, the aim is to determine its best possible weights. This study implements a two-layered neural network for modelling purposes, as shown below:

$$\sum_{i=1}^H w_i \sigma(\sum_{j=1}^d w_j \times x_j + b) \quad (17)$$

Where  $H$  describes the hidden layer neurons' quantity,  $w$  signifies the network's weightings, the bias is defined by  $b$ , and the activation function for the neurons is defined by  $\sigma$ . The main steps for the developed ENN optimization can be outlined as follows:

- 1) Set the starting positions of the population.
- 2) Compute the error function value for all of the locations. The error function is achieved as follows:

$$Error = 0.5 \times \sum_{k=1}^g \sum_{j=1}^m (Y_j(k) - E_j(k))^2 \quad (18)$$

Where  $m$  is the output nodes quantity,  $g$  stands for the training set quantity, and the output of the real and the expected output value are defined, In turn, by  $Y_j(k)$  and  $E_j(k)$ .

- 3) Save the finest place of the population.
- 4) Adjust the locations of the population by moving closer (or away from) the neighbors founded on local attractive and repulsive forces.
- 5) Introduce the random solutions movement.
- 6) Analyze the termination criteria to determine if the network has achieved the desired error rate.
- 7) Should the termination criteria be met, proceed to step 9. Otherwise:
- 8) Utilize chaos dynamics to develop chaotic positions and go to phase 2.
- 9) Stop the process

#### V. SIMULATION RESULTS

This insightful study utilizes input variables to ascertain whether breast cancer is benign or malignant accurately.

##### A. Dataset Description

The objective of this study is to leverage input variables for predicting the classification of breast cancer as either malignant or benign. To achieve this, the researchers analyzed a dataset sourced from a Wisconsin Hospital's breast cancer patient records, which was accessible via the UCI machine learning repository [20]. This dataset comprises a total of 699 examples, each characterized by nine distinct features. It's noteworthy that among these samples, sixteen instances lack complete attribute information. Collectively, these data points are represented in the form of a 683x9 matrix, where each row corresponds to a patient's data entry.

Within the dataset, a fundamental division exists between two primary categories of breast tumors: benign and malignant. The analysis reveals that out of the entire dataset, 444 patients (approximately 65%) are diagnosed with benign tumors, while the remaining 239 patients (about 35%) are diagnosed with malignant tumors. This distribution of tumor types forms a critical foundation for the subsequent analysis and model development.

The essence of the dataset lies in its nine features, which play a pivotal role in determining the classification of breast cancer. These features encompass a diverse range of parameters, including clump thickness, uniformity of cell size, and uniformity of cell shape, marginal adhesion, single epithelial cell size, bare nuclei, bland chromatin, normal



nucleoli, and mitoses. Each of these attributes holds significance in differentiating between benign and malignant tumors.

By delving into these specifics, the study provides a more comprehensive context regarding the dataset's origin, its composition, the distribution of tumor types, and the significance of the individual features that contribute to the accurate classification of breast cancer.

Table II provides the descriptive statistics of the disease and its varying range [21]. The number of network inputs is nine parameters, as in similar studies in Table II. A confusion matrix is generally used to check the success and efficiency of disease classification and diagnosis systems. The analysis of this matrix in the classification and diagnosis of patients leads to four classes of TP, TN, FN, and FP (True Positive, True Negative, False Negative, and False Positive).

From the results of the confusion matrix, three indicators of sensitivity (accuracy of the system in detecting the malignant type), specificity (accuracy of the system in detecting the benign type) and precision (the proportion of all cases that were correctly classified) are obtained, which are used to analyse the performance of the classification systems. The algorithm used in this research is presented in Fig. 2.

As the algorithm shows, in this research, after separating the test and test samples and dividing them into two benign and malignant, cancer was modelled with the help of two different types of neural networks, including the Bayesian network [22] and the standard improved ENN to demonstrate its efficiency and the outcomes were compared with each other. During the simulations, 80% of the data are employed for training, and 20% are used for testing the dataset.

### B. Results and Discussions

This research examined 685 patients with breast cancer whose data was derived from the Wisconsin hospital in the

UCI machine learning database. The number of clinical variables in each patient equalled nine risk factors. In order to train the network, 80% of the patients (548 samples) were used. The remained 10% (137) were used to test the system. As the range for each of the nine disease risk factors ranged from one to ten, Tables III to IV illustrate their frequency distributions.

Several strategies have been presented in various studies to discover the relationships between breast cancer factors. The use of artificial neural networks to diagnose the type of cancer has been investigated in various studies. In this research, we tried to use the proposed modified ENN/UIC network to improve disease diagnosis.

The confusion matrix of the proposed modified ENN/UIC network compared with the Bayesian network (BN) [22] and the standard improved ENN is shown in Fig. 3.

Fig. 3 makes it abundantly evident that, as shown by the numerous metric indicators, the proposed modified ENN/UIC model beats the modified ENN model and BN model with regard to predictive ability. Remarkably, the improved ENN/UIC model offers a breast cancer model that is more accurate. Figure 4 indicates the comparison between the three types of networks used in this research [22].

In Fig. 4, using the suggested modified ENN/UIC model with 98.54% accuracy yield the best results when compared to the other approaches, including the ENN/UIC model with 96.78% and BN model with 97.95% accuracy. The simulation results show that applying the proposed strategy yields superior outcomes when compared. The results indicated that all three types of neural networks have the ability to predict breast cancer disease with high ability.

TABLE II. INPUT VARIABLES AS MATRIX COLUMNS INCLUDING NINE FEATURES TO DETERMINE THE TYPE OF CANCER [21]

Attribute	Domain
Clump Thickness	1-10
Uniformity of Cell Size	1-10
Uniformity of Cell Shape	1-10
Marginal Adhesion	1-10
Single Epithelial Cell Size	1-10
Bare Nuclei	1-10
Bland Chromatin	1-10
Normal Nucleoli	1-10
Mitoses	1-10

TABLE III. DESCRIPTIVE STATISTICS OF THE DISEASE AND ITS VARYING RANGE

feature name	Class	Means ± SD
The thickness of the gland	Benign	2.88 ± 1.784
	malignant	7.28 ± 2.549
	Total	4.55 ± 2.932
Cell size uniformity	Benign	1.41 ± 0.967
	malignant	6.69 ± 2.835
	Total	3.26 ± 3.76
Uniformity of cell shape	Benign	1.52 ± 0.968
	malignant	6.69 ± 2.678
	Total	3.33 ± 2.899
Stickiness of edges	Benign	1.46 ± 0.928
	malignant	5.68 ± 3.288
	Total	2.94 ± 2.976
The size of a single mucous cell	Benign	2.22 ± 0.988
	malignant	5.44 ± 2.554
	Total	3.34 ± 2.334
Bare nuclei	Benign	1.46 ± 1.178
	malignant	7.74 ± 3.228
	Total	3.65 ± 3.755
Light color	Benign	2.19 ± 1.173
	malignant	5.88 ± 2.393
	Total	3.56 ± 2.561
Normal nuclei	Benign	1.37 ± 0.966
	malignant	5.97 ± 3.458
	Total	2.88 ± 3.164
Cell nucleus division	Benign	1.18 ± 0.521
	malignant	2.71 ± 2.675
	Total	1.71 ± 1.844

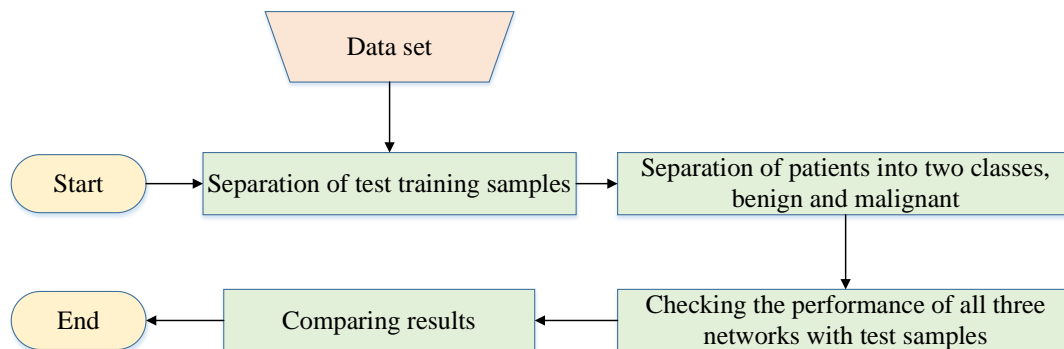


Fig. 2. The flowchart of proposed framework in this research.

TABLE IV. PREVALENCE OF NINE RISK FACTORS ASSOCIATED WITH BREAST CANCER IN REGARDED CASES

		1	2	3	4	5	6	7	8	9	10	Total
Stickiness of edges	Percent	80.6	6.0	4.8	1.8	0.9	0.5	1.4	1.2	2.1	0.0	99.3
	Abundance	571	36	32	13	5	4	10	9	15	0	695
Normal nuclei	Percent	62.4	5.3	6.1	2.7	2.8	3.2	2.4	3.4	2.2	8.7	99.2
	Abundance	423	37	43	19	20	23	17	22	16	61	681
Light color	Percent	22.6	22.8	23.1	5.7	4.8	1.4	10.3	4.2	1.7	2.8	99.4
	Abundance	151	162	163	40	35	10	72	29	12	21	695
Bare nuclei	Percent	57.6	4.4	4.1	2.8	4.4	0.7	1.2	3.1	1.4	18.8	98.5
	Abundance	403	32	29	20	31	5	9	22	10	133	694
The size of a single mucous cell	Percent	6.4	57.4	10.3	7	5.5	5.8	1.7	0.4	0.4	4	98.9
	Abundance	45	376	72	49	39	41	12	22	3	31	690
Cell nucleus division	Percent	56.3	8.4	8.4	4.8	3.4	3.0	1.8	3.7	0.6	7.8	98.2
	Abundance	394	59	59	33	24	21	12	26	4	54	686
Cell size uniformity	Percent	53.5	6.5	7.5	5.4	4.4	3.7	2.8	4.1	0.9	9.7	98.5
	Abundance	374	46	53	38	31	26	20	29	6	68	691
The thickness of the gland	Percent	19.9	7.3	14.8	11.4	18.4	4.8	3.4	6.4	2.2	9.9	98.5
	Abundance	139	51	103	80	129	34	24	45	15	69	689
Uniformity of cell shape	Percent	49.6	8.4	7.7	6.3	4.7	4.2	4.4	3.9	1.2	8.4	98.8
	Abundance	347	59	54	44	33	30	31	27	8	59	692

True class	445 (64.96%)	4 (0.58%)	99.35%	True class	438 (63.94%)	10 (1.45%)	98.28%	True class	435 (63.50%)	18 (2.62%)	97.14%
	6 (0.87%)	230 (33.58%)	96.42%		4 (0.58%)	233 (34.01%)	95.80%		14 (2.04%)	228 (33.28%)	95.37%
	98.44%	98.86%	98.54%		97.93%	98.04%	97.95%		96.44%	97.32%	96.78%
	Predicted class (A)				Predicted class (B)				Predicted class (C)		

Fig. 3. Confusion matrix of (A) modified ENN/UIC, (B) BN, and (C) modified ENN.

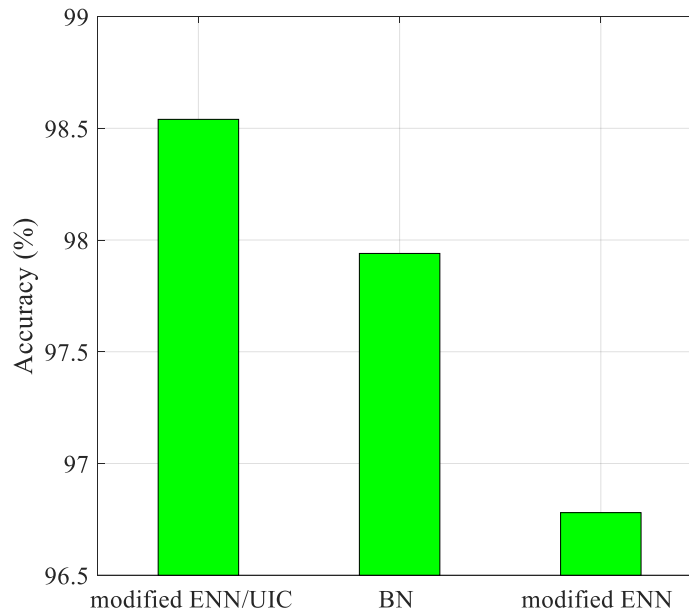


Fig. 4. Comparison between three types of networks used in this research.

## VI. CONCLUSION

Among different cancers, one cancer that starts from breast cells is breast cancer. Both men and women are affected by this disease, but this cancer is more prevalent among women. Symptoms of breast cancer can consist of a mass in the breast, a change in the form or measure of the breast, variations to the skin over the breast, like puckering or dimpling, and a discharge from the tit. Breast cancer treatment can be various, contingent on the sort and phase of the disease, but may comprise chemotherapy, radiation therapy, surgery, hormone therapy, and aimed therapy. Many deep neural networks show excellent results in accurately classifying tumor cells. Therefore, a deep artificial neural network can be used along with other non-invasive diagnostic methods that are usually used (such as mammography and radiography), as a diagnostic support system with high sensitivity and specificity to identify benign and malignant breast tumors. The existing study aimed to diagnose benign and malignant breast cancer using a new method based on deep neural networks and an improved meta-heuristic algorithm. Since artificial neural networks based on

deep calculations in the diagnosis of diseases have attracted the attention of many researchers in recent years, therefore, in this research, a novel optimized and improved process was used for classifying the type of breast cancer into two categories, malignant and benign. First, a modified version of the Elman Neural Network was trained with samples. Then, it was optimized on the basis of an upgraded version of the imperialist competitive algorithm.

At last, the model was evaluated with test samples. The model validation was performed based on a breast cancer dataset of the Wisconsin hospital. For a failed comparison, the outcomes of the proposed technique were validated by two other methods, including the Bayesian network and the standard improved ENN without optimization. Outcomes showed that all three methods provide good accuracy for a cancer diagnosis; using the proposed method for this aim achieved higher accuracy.

Looking forward, advancements in nanotechnology hold the potential to revolutionize early detection and monitoring of

breast cancer. Researchers are exploring the development of highly sensitive and specific nanosensors that can detect subtle molecular changes associated with cancer cells. These nanosensors could be integrated into wearable devices, providing continuous monitoring of biomarkers in bodily fluids. This real-time data could offer clinicians valuable insights into disease progression and treatment response, enabling more precise and timely interventions. Moreover, the combination of nanotechnology with AI could enhance the accuracy of data interpretation, leading to earlier and more accurate diagnoses. While this field is still in its infancy, the convergence of nanotechnology and AI could pave the way for groundbreaking advancements in breast cancer management, ultimately improving patient outcomes and quality of life.

#### REFERENCES

- [1] S.L. Jacob, L.A. Huppert, H.S. Rugo, Role of immunotherapy in breast cancer, *JCO Oncology Practice*, p. OP. 22.00483, 2023.
- [2] R.C. Chan, C.K.C. To, K.C.T. Cheng, T. Yoshikazu, L.L.A. Yan, G.M. Tse, Artificial intelligence in breast cancer histopathology, *Histopathology*, 82, pp. 198-210, 2023.
- [3] E. Castro, J.C. Pereira, J.S. Cardoso, Symmetry-based regularization in deep breast cancer screening, *Medical Image Analysis*, 83, p. 102690, 2023.
- [4] B. Gyawali, M. Bowman, I. Sharpe, M. Jalink, S. Srivastava, D.T. Wijeratne, A Systematic Review of eHealth Technologies for Breast Cancer Supportive Care, *Cancer Treatment Reviews*, p. 102519, 2023.
- [5] R. Wu, J. Luo, H. Wan, H. Zhang, Y. Yuan, H. Hu, J. Feng, J. Wen, Y. Wang, J. Li, Evaluation of machine learning algorithms for the prognosis of breast cancer from the Surveillance, Epidemiology, and End Results database, *Plos one*, 18, p. e0280340, 2023.
- [6] F. Hamedani-KarAzmoodehFar, R. Tavakkoli-Moghaddam, A.R. Tajally, S.S. Aria, Breast cancer classification by a new approach to assessing deep neural network-based uncertainty quantification methods, *Biomedical Signal Processing and Control*, 79, p. 104057, 2023.
- [7] K. Vijayakumar, V.J. Kadam, S.K. Sharma, Breast cancer diagnosis using multiple activation deep neural network, *Concurrent Engineering*, 29, pp. 275-284, 2021.
- [8] Q.A. Al-Hajja, A. Adebajo, Breast cancer diagnosis in histopathological images using ResNet-50 convolutional neural network, 2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), IEEE, pp. 1-7, 2020.
- [9] S. Bourouis, S.S. Band, A. Mosavi, S. Agrawal, M. Hamdi, Meta-heuristic algorithm-tuned neural network for breast cancer diagnosis using ultrasound images, *Frontiers in Oncology*, 12, p. 834028, 2022.
- [10] Y.-c. Cheng, W.-M. Qi, W.-Y. Cai, Dynamic properties of Elman and modified Elman neural network, *Proceedings. International Conference on Machine Learning and Cybernetics*, IEEE, pp. 637-640, 2002.
- [11] W. Jia, D. Zhao, Y. Zheng, S. Hou, A novel optimized GA-Elman neural network algorithm, *Neural Computing and Applications*, 31, pp. 449-459, 2019.
- [12] Q. Wu, Q. Zhu, S. Han, Elman Neural Network-Based Direct Lift Automatic Carrier Landing Nonsingular Terminal Sliding Mode Fault-Tolerant Control System Design, *Computational Intelligence and Neuroscience*, 2023, 2023.
- [13] G. Ren, Y. Cao, S. Wen, T. Huang, Z. Zeng, A modified Elman neural network with a new learning rate scheme, *Neurocomputing*, 286, pp. 11-18, 2018.
- [14] E. Atashpaz-Gargari, C. Lucas, Imperialist competitive algorithm: an algorithm for optimization inspired by imperialistic competition, 2007 IEEE congress on evolutionary computation, Ieee, pp. 4661-4667, 2007.
- [15] S. Talatahari, B.F. Azar, R. Sheikholeslami, A. Gandomi, Imperialist competitive algorithm combined with chaos for global optimization, *Communications in Nonlinear Science and Numerical Simulation*, 17, pp. 1312-1319, 2012.
- [16] D. Lei, Y. Yuan, J. Cai, D. Bai, An imperialist competitive algorithm with memory for distributed unrelated parallel machines scheduling, *International Journal of Production Research*, 58, pp. 597-614, 2020.
- [17] S.M.G. Kashikolaie, A.A.R. Hosseinabadi, B. Saemi, M.B. Shareh, A.K. Sangaiah, G.-B. Bian, An enhancement of task scheduling in cloud computing based on imperialist competitive algorithm and firefly algorithm, *The Journal of Supercomputing*, 76, pp. 6302-6329, 2020.
- [18] Y. Tang, F. Zhou, An improved imperialist competition algorithm with adaptive differential mutation assimilation strategy for function optimization, *Expert Systems with Applications*, 211, p. 118686, 2023.
- [19] N. Razmjoo, M. Ramezani, N. Ghadimi, Imperialist competitive algorithm-based optimization of neuro-fuzzy system parameters for automatic red-eye removal, *International Journal of Fuzzy Systems*, 19, pp. 1144-1156, 2017.
- [20] Breast Cancer Wisconsin (Diagnostic) Data Set, in: U.M.L. Repository (Ed.), 2017.
- [21] American Cancer Society Information and Resources about for Cancer: Breast, Colon, Lung, Prostate, Skin, 2020.
- [22] H. Chen, W. Lu, Y. Zhang, X. Zhu, J. Zhou, Y. Chen, A Bayesian network meta-analysis of the efficacy of targeted therapies and chemotherapy for treatment of triple-negative breast cancer, *Cancer medicine*, 8, pp. 383-399, 2019.

# Stacked LSTM and Kernel-PCA-based Ensemble Learning for Cardiac Arrhythmia Classification

Azween Abdullah<sup>1</sup>, S. Nithya<sup>2</sup>, M. Mary Shanthi Rani<sup>3</sup>, S. Vijayalakshmi<sup>4</sup>, Balamurugan Balusamy<sup>5</sup>

Faculty of Applied Science and Technology, Pedana University, Kuala Lumpur, Malaysia<sup>1</sup>

The Gandhigram Rural Institute (Deemed to be University), Gandhigram<sup>2,3</sup>

Christ University, Pune<sup>4</sup>

Shiv Nadar Institution of Eminence<sup>5</sup>

Center for Global Health Research, Saveetha Medical College and Hospital,

Saveetha Institute of Medical and Technical Sciences, India<sup>5</sup>

**Abstract**—Cardiovascular diseases (CVD) are the most prevalent causes of death and disability worldwide. Cardiac arrhythmia is one of the chronic cardiovascular diseases that create panic in human life. Early diagnosis aids physicians in securing life. ECG is a non-stationary physiological signal representing the heart's electrical activity. Automated tools to detect arrhythmia from ECG signals are possible with Machine Learning (ML). The ensemble learning technique combines the power of two or more classifiers to solve a computational intelligence problem. It enhances the performance of the models by fusing two or more models, which extremely increases its strength. The proposed ensemble Machine learning amalgamates the potency of Long Short-Term Memory (LSTM) and ensemble learning, opening up a new direction for research. In this research work, two novel ensemble methods of Extreme Gradient Boosting-LSTM (EXGB-LSTM) are developed, which use LSTM as a base learner and are transformed into an ensemble learner by coalescing with Extreme Gradient Boosting. Kernel Principal Component Analysis (K-PCA) is a significant non-linear dimensionality reduction technique. It can manage high-dimensional datasets with various features by lowering the dimensionality of the data while retaining the most crucial details. It has been applied as a preprocessing step for feature reduction in the dataset, and the performance of EXGB-LSTM is tested with and without K-PCA. Experimental results showed that the first method, fusion of EXG-LSTM, has reached an accuracy of 92.1%, Precision of 90.6%, F1-score of 94%, and Recall of 92.7%. The second proposed method, KPCA with EXGB-LSTM, attained the highest accuracy of 94.3%, with a precision of 92%, F1-score of 98%, and Recall of 94.9% for multi-class cardiac arrhythmia classification.

**Keywords**—Arrhythmia classification; ensemble learning; extreme gradient boosting; kernel PCA; LSTM; machine learning

## I. INTRODUCTION

Cardiovascular diseases have become a significant public health problem globally, accounting for an average of 30% of all deaths worldwide in 2008. With the rapid increase in the world population, cardiovascular disease prevalence has been on a steady rise, increasing related mortality and morbidity. Recent statistical data indicate that CVD caused 17 million fatalities in 2015 [1].

This rising trend in cardiovascular diseases presents a challenge to both developed and developing countries.

Moreover, they cause a major economic burden on nations due to the significant costs associated with treatment and management [2].

Each patient's clinical examinations and medical histories provide the basis of the conventional CVD diagnosis paradigm. These results are evaluated under a set of quantitative medical indicators to categorize the patients based on the taxonomy of medical disorders [3].

One of the most dangerous symptoms of heart disease is cardiac arrhythmia. Early arrhythmia prognosis is becoming more significant in heart failure research since it can support clinical patient treatment. Diagnosing potentially fatal cardiac arrhythmias now requires the competence of cardiologists [4].

Cardiac Arrhythmia is an abnormal phenomenon observed in the electrical activity of the human heart. The irregular heart rhythm indicates numerous complications in the heart chamber. This condition is due to an unhealthy lifestyle and various medical conditions [5]. Bradycardia and tachycardia are the two main categories of arrhythmias. Bradycardia is an arrhythmia in which the heartbeat is less than 60 beats per minute (bpm), while tachycardia is an arrhythmia in which the heartbeat can increase to 100 bpm [6]. A sudden onset of a cardiac problem, such as inadequate heart blood flow, shortness of breath, chest pain, fatigue, or unconsciousness, is how arrhythmia often occurs. Abnormal ECG indicates the symptoms. As a result, it's critical to recognize and address arrhythmia as soon as possible [7]. Cardiologists typically spend a lot of time classifying arrhythmias with great precision. Automatic arrhythmia classifiers based on AI algorithms can assist cardiologists in getting higher precision and spending less time diagnosing patients [8].

The Electrocardiogram (ECG) is a painless method to register the electrical impulse generated by the heart in a human being. It is a time-varying signal that reflects ionic current flow resulting from the contraction and subsequent relaxation of the heart [9]. ECG is composed of the P wave that corresponds to the electrical activity of the atria, the Q, R, and S waves composing the QRS complex that corresponds to depolarization of the ventricles, and the T wave that registers the repolarization of the ventricles [10]. Furthermore, the R-R intervals, the time difference between two R waves, can be

obtained from the QRS complex. The different components of the ECG signal have specific shapes in the time and frequency domains, which allow specialists to detect possible anomalies by analyzing the ECG signal in these two domains. Moreover, the ECG signal is stochastic and can include abnormal components such as narrower or wider QRS complexes. The abnormal ECG signal can provide important information about cardiac abnormalities and may aid in diagnosing various heart conditions. However, there is still much to learn about the ECG signal and its uses in clinical settings. In recent years, researchers have made significant progress in enhancing the accuracy and efficacy of ECG signals [11].

The ECG signal is a crucial diagnostic tool for identifying heart diseases and assessing cardiac health, so understanding its properties and analysis techniques is paramount. The abnormal heart rhythms are recorded through ECG to help the physicians identify the abnormal rhythm. Earlier diagnosis of irregular beats can avoid various cardiac complications. Prediction of cardiovascular disease is a tedious process for the cardiologist. Hence, the automated arrhythmia diagnosis will assist clinicians in diagnosing and providing timely patient treatment, averting fatalities [12].

The situation will worsen in areas with inadequate medical professionals and clinical equipment, particularly in emerging nations. This drives the need for a trustworthy, automated, affordable monitoring and diagnosing system. The need for this is growing among healthcare professionals so that using computer-aided diagnosis (CADS) systems can be linked to doing the necessary medical assessments [13].

ML is a powerful tool playing a pivotal role in every field, such as Healthcare, Text classification, Plant disease classification, Robotics, Compression, Banking, etc., [14 - 20].

With advancements in technology and signal processing techniques, the accuracy of ECG signal analysis has significantly improved. Furthermore, using ML algorithms has shown promising results in enhancing accuracy in diagnosing arrhythmia.

The significance of effective and precise arrhythmia categorization and detection is becoming more evident with the advent of remotely controlled healthcare systems for heart disease patients. Over the past several years, various ML techniques have been built into diagnostic systems to help with the difficult challenge of accurately classifying arrhythmias from recorded ECG data. Selecting the best methods for identifying and categorizing cardiac disease might be challenging. It entails considering context, analyzing data, and the needs of particular patients' details [21].

Effective ML techniques enable implicit programming-free ML. The goal is to create an algorithm that can take a set of patterns and automatically generalize from preliminary data with or without human intervention. Cluster analysis, or clustering, is a component of unsupervised ML techniques [22-26].

Recently, boosting algorithms have been applied to accelerate the performance of ML algorithms. The concrete objective of this research work is to develop ensemble learning for predicting cardiac arrhythmia challenges.

The novelty of the article includes the following significant contributions:

- Development of Ensemble Model for effective Arrhythmia classification.
- Implementation of fusion of Stacked LSTM and XGBoost algorithms in the classification.
- Application of K-PCA for Feature Reduction.
- Identification of 16 types of arrhythmias with enhanced accuracy by hyper-parameter tuning.
- Validation and comparison of the performance of the proposed models with state-of-the-art methods.

The organization of the paper is outlined as follows. Section I presents the introduction. Section II elaborates on the literature survey in related fields and datasets. Section III presents a brief note on the dataset used for experimentation, followed by the methods used for classification. Section IV depicts the experiments and results obtained, followed by a discussion. Finally, the work is concluded in Section V.

## II. LITERATURE REVIEW

Numerous methods are currently available for classifying heart arrhythmias. Some techniques use ML techniques, including support vector machines, neural networks, decision trees, and K Means. This section presents a brief survey of ML and DL-based methods for arrhythmia classification.

Ozcift et al. focused on the following two strategies: i) To choose pertinent characteristics from the cardiac arrhythmia dataset, using a correlation-based feature selection algorithm. ii) The effectiveness of the suggested training method is assessed using the performance of selected features with and without simple random sampling using the Random Forest (RF) ML algorithm. This method achieved an accuracy of 90% [27].

Namsrai et al. developed a new ensemble method using feature selection schema to derive several feature subsets from the original dataset. The feature subsets are then used to train classification models, and the classification performance of each feature set is combined with a voting approach. The ensemble classifier reportedly beats the classifiers learned on the original dataset by 7.28 % for Naive Bayes (NB), 2.27 % for the Support Vector Machine (SVM), 17.84 % for the Decision Tree (DT), and 13.43 % for the Bayes Network of F-score [28].

Guvendir et al. presented a Voting Feature Intervals (VFI), a supervised ML algorithm for detecting arrhythmia classification. A majority vote among the class predictions produced by each characteristic individually is the basis for how it operates. The VFI algorithm performs better than other standard algorithms like Naive Bayesian and K-Nearest Neighbour (KNN), according to comparisons [29].

Freund et al. proposed a novel Adaptive Boosting (AdaBoost) algorithm and the ensemble technique based on the Adaptive Boosting algorithm by engrafting the decision tree into a framework. They created the new boosting algorithm by

using the multiplicative weight update technique. This boosting approach doesn't need any prior understanding of how well the weak learning algorithm performs. Additionally, they investigated how the novel boosting approach can be applied to problems involving learning functions whose domain, rather than being binary, is an arbitrarily finite set or a bounded segment of the real line [30].

Friedman et al. developed a general gradient descent boosting paradigm for additive expansions based on any fitting criterion. Tools for analyzing Tree Boost models are described, and special enhancements are built for the particular case where the individual additive components are regression trees. Gradient boosting of regression trees generates highly competitive, dependable, and comprehensible regression and classification algorithms that are particularly suitable for mining imperfect data [31].

Khemchandani et al. developed Twin SVM (TSVM) since conventional SVM has computational complexity. The developed twin SVM is faster in computation and has good generalization for binary classification [32].

Dhyani et al. employed SVM and the 3D Discrete Wavelet Transform (DWT) to characterize and analyze ECG signals. SVM is used to classify the ECG using the nine types of heartbeats identified by the different classifiers. In comparison to the complex support vector machine (CSVM) 98.5% and weighted support vector machine (WSVM) 99%, the SVM classifier has a normal precision of 99.02% [33].

Mohanty et al. experimented with the Cubic Support Vector Machine (CSVM) and the C4.5 classifiers for the classification of Ventricular Fibrillation (VF), ventricular tachycardia (VT), and normal sinus rhythm (NSR). Early detection and prevention of ventricular arrhythmias require the ability to predict VT and ventricular fibrillation VF. By combining temporal, spectral, and statistical data, the researchers described a method for identifying and categorizing VT and VF arrhythmias. The results of this study indicate that the suggested approach outperformed cubic SVM, which had an accuracy of 92.23%, with a sensitivity of 90.97%, specificity of 97.86%, and accuracy of 97.02% in the C4.5 classifier [34].

Yadav et al. determined the significance of particular elements pertinent to diagnosing cardiac arrhythmia. They employed a random forest classifier model, measuring the significance of many features like blood pressure, sodium, potassium, calcium, and respiratory rate, among other features. Hyperparameter optimization methods like grid search and genetic algorithms are compared to find the maximum number and depth of trees in the forest. Area Under the Receiver Operator Curve (AUC) score of 0.9787 was the model's highest performance [35].

Kumari et al. researched classifying arrhythmias more accurately and quickly than before. The research uses an SVM classifier using Discrete Wavelet Transform (DWT). DWT was used to extract 190 features from the prepared data. The SVM classifier, which is the best for classification, was given the retrieved features. The testing set was used to assess the results,

and a confusion matrix was used to plot the final results. The model's performance accuracy is 95.92 % [36].

Taher et al. tested with a fusion-based model, which is used to choose the highest-ranked features most effectively. The fusion's output is subsequently subjected to an ensemble of classifiers. High-ranked features acquired from several heuristic feature-selection algorithms are fused using a fuzzy-based feature-selection fusion method. The three primary sections of the work are sensing data and preprocessing, feature queuing, selection and extraction from features, and the predictive model. The method performs classification tasks with higher accuracy, F1 measure, recall, and precision. For binary and categorized class modes, respectively, it obtains an accuracy of 98.5% and 98.9% [37].

Ayar et al. suggested a model for binary and multi-class classification, and the model used genetic algorithms for feature selection and the C4.5 algorithm with DT for classification. The average accuracy for binary classification was 86.96%, and for multi-class classification, it was 78.76% [38].

Jadhav et al. proposed an approach based on feature elimination to diagnose arrhythmia in binary classification. With ensemble sizes of 15 and 20, the random subspace ensemble classifier can identify arrhythmias with an accuracy of 91.11% [39].

Khan et al. (2018) depicted a two-stage cascade structure technique for categorizing cardiac arrhythmia. The first stage used logistic regression (LR), DT, and RF to classify arrhythmias. In contrast, the second stage used Multi-Layer Perceptron's (MLP) and then LSTM to classify arrhythmias into many classes and binary classes. The accuracy obtained by the first experiment for the LR, DT, and RF was 83.0%, 85.4%, and 87.5%, respectively. The accuracy obtained for the MLP and LSTM experiments was 89.0% and 94.8%, respectively [40].

Itzhak et al. proposed a study on detecting three forms of arrhythmia: Atrial fibrillation, bradycardia, and tachycardia. The authors used random forest, SVM, and logistic regression as classifiers. Their best result was a multiclass random forest classifier that performed with a sensitivity and specificity of 0.92 and 0.86, respectively [41].

Chang et al. found 12 different cardiac rhythm classes using an LSTM model. The LSTM model had an accuracy of 0.982 and an AUC of 0.987 for classifying each of the 12 cardiac rhythms. The F1 score was 0.777, and the precision and recall were 0.692 and 0.625, respectively [42].

Xu et al. developed an innovative ML approach called the Twin Support Vector Machine (TSVM), which seeks to identify two nonparallel planes for each class. The binary classification issue can be solved using traditional TSVM. The multi-class categorization problem can be solved with the Twin-KSVC multi-class classification technique. It combines the benefits of the Support Vector Classification Regression Machine for K-class classification (K-SVCR) with both TSVM and K-SVCR [43].

Mustaqeem et al. selected the wrapper feature selection method in conjunction with SVM-based methods like one-against-one (OAO), one-against-all (OAA), and an error-correction code (ECC) for determining whether an arrhythmia is present or absent. Kappa statistics, accuracy, and Root Mean Square Error (RMSE) measures are used to measure the performance. OAO outperformed all others by scoring 92.07% [44].

Khan et al. tested with Principal Component Analysis (PCA) for feature reduction technique in classifying arrhythmia. The Deep Learning (DL) architecture LSTM is used for classification. An accuracy of 93.5% was attained using PCA and LSTM together [45].

Chen G et al. suggested a wrapper method called Cosine Similarity Measure Support Vector Machines (CSMSVM) by introducing the cosine distance into SVM to reduce features. Traditionally, feature selection methods have extracted features and learned SVM parameters separately or in the attribute space. This may have led to a loss of information about the classification process or increased classification error when the kernel SVM was included. The suggested CSMSVM framework combines feature selection, SVM parameter learning, and removing low-relevance features to achieve better performance [46].

Chandrashekar et al. experimented with several feature selection methods since each underlying algorithm will respond differently to different data types. Using feature selection approaches demonstrates that more information may not be better in ML applications. The feature selection technique for the application is chosen based on the following factors: simplicity, stability, number of reduced features, classification accuracy, storage needs, and computing requirements. It will have advantages like stronger classifier models, improved generalization, and identification of irrelevant variables [47].

Celin et al. used filtering techniques such as low pass, high pass, and Butterworth filter to eliminate the high-frequency noises. The peak detection algorithm is used to locate the peaks, and statistical parameters extract the signal's features. The ML classifiers SVM, Adaboost, ANN, and NB were implemented. The accuracy of the SVM, Adaboost, ANN, and NB classifiers is 87.5%, 93%, 94, and 99.7%, respectively, according to experimental results [48].

Park. J et al. suggested a novel heartbeat classification model that combines an adaptive feature extraction approach and cascade classifiers. The cascade classifiers are constructed from two different random forest classifiers. This can be used with normalized beat morphology features in an environment with limited resources. This model's accuracy was 97.34 %. The random forest classifier was chosen to reduce computational costs and memory even though the k-NN classifier performed better on the initial feature set [49].

Amorim et al. implemented Contourlet, and Shearlet transforms to separate ECG signals into different frequency bands and then extracted features from time-frequency coefficients. They evaluated the performance of KNN, SVM, and Random Forest classifiers on these collected features to

categorize seven different forms of beat arrhythmia. With an accuracy of 91.32% and a sensitivity of 90.23%, they used features based on contourlet transforms to get the greatest performance for random forests [50].

Romdhane et al. presented a CNN model to design an algorithm for heartbeat segmentation. This approach is straightforward and efficient because it doesn't need any signal processing that depends on prior knowledge of the signal's morphology or spectrum. The work focused on imbalanced datasets and designed a deep CNN model using a novel focal loss function. The evaluation's findings showed that the targeted loss function increased the overall metrics and the classification accuracy for imbalanced classes. Our suggested strategy achieved 98.41% accuracy, 98.38% F1-score, 98.37% precision, and 98.41% recall [51].

Sharma et al. employed both noisy and denoised (clean) ECG signals. The ECG signals were wavelet decomposed using a dyadic orthogonal filter bank with stop-band energy (SBE) minimization. Characteristics based on fuzzy entropy, Renyi entropy, and fractal dimension were then extracted for a quicker and more precise classification into the five classes. These features were then fed into classifiers, with the KNN classifier attaining maximum accuracy (MAAC) of 98.1%, maximum sensitivity (MASE) of 85.63%, and maximum specificity (MASP) of 98.27% for clean data and MAAC of 98.1%, maximum sensitivity (MASE) of 85.33% for noisy data. This technology can be utilized in intensive care units to help clinicians make fast diagnoses [52].

Chen et al. developed a novel methodology of piecewise linear splines for feature selection and a gradient-boosting algorithm to classify atrial fibrillation and other cardiac dysrhythmias. A piecewise linear spline is used to fit the ECG waveform, and morphological properties associated with the coefficients of the piecewise linear spline are retrieved. The morphological coefficients and heart rate variability features are categorized using XGBoost. The algorithm received an average F1 score on the independent testing set of 81% and an F1 score on a 10-fold cross-validation of 81%. With certain morphological features, the system performs well on multi-label short ECG classification [53].

Assodiky et al. employed LSTM to categorize the types of arrhythmias, using AdaDelta as the adaptive learning rate method and improved performance. It was contrasted with LSTM, which lacked an adaptive learning rate. The best outcome that demonstrated great accuracy was produced by combining LSTM and AdaDelta. The accurate classification rate for train and test data was 98% and 97%, respectively [54].

Although the aforementioned techniques have shown some reasonable accuracy in classifying arrhythmias, more work has to be done, such as increasing accuracy. Most arrhythmia classification techniques currently in use are based on traditional ML techniques, which emphasize careful feature selection and classification less. DL techniques have found widespread usage due to the quick development of tools for working with high-dimensional data and advancements in computing power. Despite several research works reported in the literature, there is still a pressing need for novel strategies



to improve the accuracy and effectiveness of classification models for Arrhythmia.

### III. MATERIALS AND METHODS

#### A. Experimented Dataset

The proposed research work has been experimented with the Arrhythmia dataset available in the UCI Irvine ML repository. The dataset contains data on 452 patients in total, and each record has 279 features that describe the traits of an arrhythmia. Age, sex, height, and weight are the first four characteristics that best describe the subject. The remaining 274 features are taken from recordings of typical 12-lead ECGs. The heart muscle depolarizes during each heartbeat, causing small electrical changes on the skin that are detected and amplified by the ECG. A 12-lead ECG captures 12 distinct electrical heart impulses at approximately the same time. Arrhythmia data has an extensive feature size, which may need a feature reduction approach. It has numerous missing values. Some columns with missing values are eliminated, and the experiments are carried out. The goal is to categorize cardiac arrhythmia into one of the three categories and to distinguish between its presence and absence. Class 01 denotes regular ECG classes, Class 02 through Class 15 denotes various arrhythmia classes, and Class 16 denotes the remaining unclassified ones [63].

#### B. Methods

The proposed work reports on two methodologies EXGB-LSTM and KPCA-EXG-LSTM. The proposed methods classify arrhythmia through a pipeline that includes extreme boosting and stacked LSTM together and implemented in the preprocessed dataset. The flow diagram of the proposed work is shown in Fig.1.

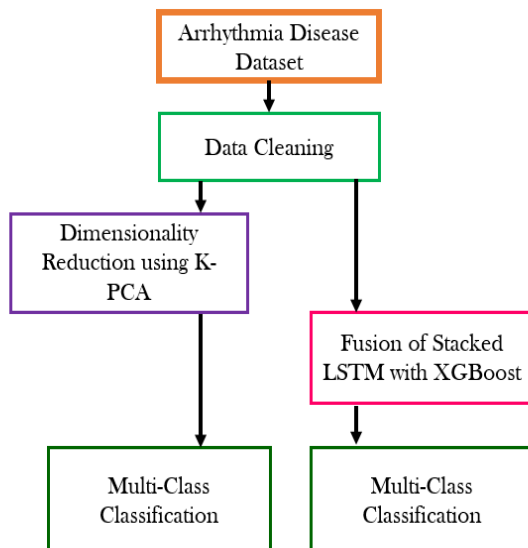


Fig. 1. A flow diagram of EXG-LSTM with K-PCA.

The implementation of the proposed model EXGB-LSTM is shown in the Algorithm 1.

<b>Algorithm 1 EXGB-LSTM</b>
<b>Input:</b> UCI Arrhythmia Machine Learning Dataset
<b>Output:</b> Classification of 16 classes of Arrhythmia
<b>Begin</b>
<b>Phase I : Data Cleaning</b>
Eliminating columns with more number of missing values.
<b>Phase II : Applying Stacked LSTM with XGBoost</b>
Fusion of Stacked LSTM with XGBoost in the Processed dataset.
<b>Phase III: Multi-Class Classification of Arrhythmia</b>
<b>End</b>

In the second method, the dataset is dimensionally reduced by K-PCA, followed by ensemble learning using LSTM and extreme boosting and classification. The steps of the proposed K-PCA with the EXGB-LSTM model are outlined below in the Algorithm 2.

<b>Algorithm 2 K-PCA + EXGB-LSTM</b>
<b>Input:</b> UCI Arrhythmia Machine Learning Dataset
<b>Output:</b> Classification of 16 classes of Arrhythmia
<b>Begin</b>
<b>Phase I : Data Cleaning</b>
Eliminating columns with more missing values.
<b>Phase II : Dimensionality Reduction using Kernel-PCA</b>
<b>Phase III: Applying Stacked LSTM with XGBoost</b>
Fusion of Stacked LSTM with XGBoost in the Processed dataset.
<b>Phase IV : Multi-Class Classification of Arrhythmia</b>
<b>End</b>

#### C. Kernel Methods

Dimensionality reduction plays a significant role in effectively handling massive dimensional data. The motive of dimensionality reduction may be noise reduction, pre-processing, compression, etc. PCA is a mathematical approach to converting several correlated variables into uncorrelated variables known as Principal components (PC). This PC represents the maximum variance in the dataset. PCA works only for linear structures, and K-PCA has been developed as a non-linear extension of standard PCA. The primary notion of K-PCA is to map the original data into a huge dimensional space through a specific function and then implement the PCA algorithm to it. The linear PCA of the vast dimensional feature

space represents a non-linear. Most interesting directions can be found in the original input space by the PCA [64]. It is mainly applied in complex non-linear data such as 3D reconstruction, handwritten digits, bioinformatics face images, natural signals, and geostatistics kriging. Various kernel methods include polynomial, Gaussian, Laplace Radial Basis Function, Sigmoid, and hyperbolic tangent. One of the merits of K-PCA is that original data can be regenerated from its principal components [65] in the following three steps.

- Computing Covariance Matrix

The Covariance Matrix(C) is used to represent the relationship between two random variables and is defined in Eq. (1):

$$C = \frac{1}{n} \sum_{i=1}^n \phi x^i \phi x_i^T \quad (1)$$

where  $\phi x^i$  represents the input feature space,  $x^T$  represents the Transpose and  $x^i$ , is one of the 'n' multivariate observations. Also, the mean of the data in the feature space is assumed to be zero.

- Decomposition of Eigenvalues

Eigen decomposition of an input data distribution results in the generation of eigenvectors, which form the principal components of the data as given in Eq. (2):

$$\lambda v = Cv \quad (2)$$

where 'λ' is the eigenvalue and 'v' is the eigenvector which is a linear combination of samples.

- Selection of Eigenvectors and transform data

A function that holds the vectors in the original input space and when the dot product of the vectors in the feature space is returned then it is known as a kernel function as is shown in Eq.(3):

$$K(x_i, x_k) = \phi x_i^T \phi x_k \quad (3)$$

Kernel extracts up to 'n' samples in non-linear PCs without expensive computations [65].

#### D. LSTM

A conventional Recurrent Neural Network (RNN) is an extension of a feedforward neural network that can handle variable lengths of sequential input, also called an Auto Associative or Feedback Network. The neural network model is structured to preserve previously hidden neurons' output, and it is fed as input to the next cell [66]. LSTM was first proposed in 1997 by Sepp Hochreiter and Jurgen Schmidhuber. A modified version of RNN makes each recurrent unit adaptively capture dependencies of different time scales. It has a forget cell to modulate the flow of information. RNN suffers from complex training time for longer sequences and vanishing gradient problems. LSTM was structured to eliminate gradient problems with three gates namely, an input gate  $i_t$ , an output gate  $o_t$ , and forget gate. . The design of the LSTM cell is represented in Fig. 2.

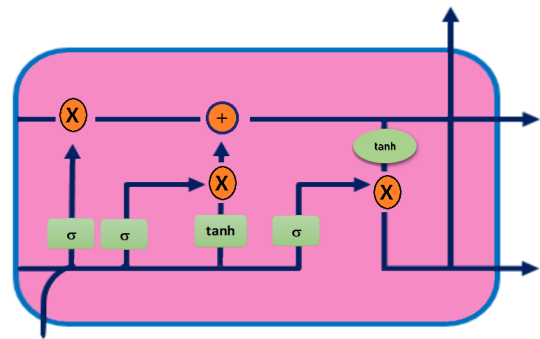


Fig. 2. LSTM cell.

The Model of LSTM is working in principle by representing the input sequences as  $\{x_1, x_2, x_3, \dots, x_t\}$ ,  $\{y_1, y_2, y_3, \dots, y_t\}$  as an output sequence and  $\{S_1, S_2, S_3, \dots, S_t\}$  as an internal sequence of cell states.

In the LSTM cell, the three gates are computed with Eq.(4-5):

$$i_t = \sigma (W_i \cdot [y_{t-1}, x_t] + b_i) \quad (4)$$

$$f_t = \sigma (W_f \cdot [y_{t-1}, x_t] + b_f) \quad (5)$$

$$O_t = \sigma (W_o \cdot [y_{t-1}, x_t] + b_o) \quad (6)$$

where  $W_i$ ,  $W_f$ , and  $W_o$  denote the weight of the corresponding input, and  $b_i$ ,  $b_f$ , and  $b_o$  are the respective bias in the cell. The activation function sigmoid is represented as 'σ' in the equation.

The candidate state  $\hat{S}_t$  is computed with the Eq. (7):

$$\hat{S}_t = \tanh (W_s \cdot [y_{t-1}, x_t] + b_s) \quad (7)$$

The current cell state value  $S_t$  is calculated from the previous cell state value  $S_{t-1}$  using Eq. (8):

$$S_t = f_t * S_{t-1} + i_t * \hat{S}_t \quad (8)$$

The output of the cell  $y_t$  is determined by Eq. (9):

$$y_t = o_t * \tanh(S_t) \quad (9)$$

The input gate in the LSTM model chose what to store in the current cell. The output gate determines what is to be passed over. The forget gate determines how to forget the current state [65].

#### E. Stacked LSTM

Multiple hidden layers in stacked LSTM architectures can be used to increase the amount of abstraction gradually. They operate better because they give a more accurate representation of sequence data. In the stacked LSTM architecture, the input from the previous LSTM hidden layer is passed into the current LSTM hidden layer. This has the potential to improve neural network performance significantly. Including enough layers in the model improves the LSTM model's accuracy and dependability. Fewer neurons are needed in each layer, which cuts down on training time. In this study, a stacked LSTM network is suggested with the fusion of XGBoost [67, 68]

F. Extreme Gradient Boosting

The eXtreme Gradient Boost (XGBoost) is a decision tree-based algorithm that transforms weak learners into strong learners. Boosting is an iterative algorithm compared to bagging, a parallel algorithm. The computational procedure for constructing a decision tree in gradient boosting is very time-consuming. XGBoost is an optimized version of the gradient boosting algorithm, which improves the training time. Many models were trained on different subsets of data, and the best-performing model was chosen. It is best suited for non-linearity problems [69, 70]. Combining the models is part of ensemble learning, where different models are combined to optimize the algorithm's performance. Stacking is one of the usual processes in combining models and training an ensemble model on the classification, enabling convenient experimentation and model comparisons. Even though Light Gradient Boosting is faster when compared with the performance it is equivalent [71-73]. The framework of the developed method is presented in Fig. 3.

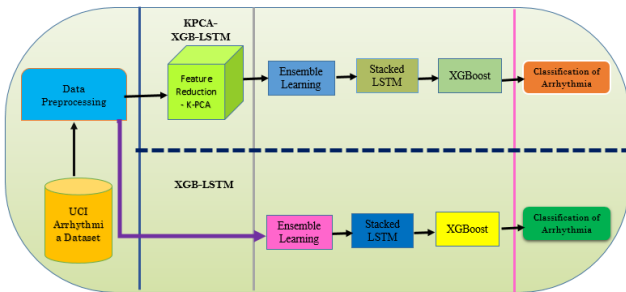


Fig. 3. A framework of EXG-LSTM with K-PCA.

IV. EXPERIMENTATION AND ANALYSIS

A. Experimental Setup

The proposed model uses an Intel Core(R) i5-9300H CPU running at 2.40 GHz, an NVIDIA GeForce® GTX 1050 graphics card, a 64-bit operating system, and 8GB of RAM.

The experiment on ensemble LSTM was carried out with the Arrhythmia dataset. The medical dataset consists of 206 linear valued features and a few categorical features among the 279 features. 452 instances are present in the dataset. It has been distinguished into 16 groups to indicate the existence and non-existence of arrhythmia. Class 1 shows healthy or normal samples. The total number of healthy samples is 245. Classes 2-15 are categorized as different types of arrhythmias. Class 16 indicates the other unclassified arrhythmia. As the dataset may have null values for certain features. Those features having more missing values are removed during the pre-processing stage. The remaining features with fewer missing values are treated by the statistical mean method. The proposed work was tried in two different ways. The dataset was chosen directly after applying data cleaning and data normalization methods to implement LSTM in the first experiment. To improve the performance of LSTM, XGBoost was added, and it showed a better performance. In the second method, the dataset was further subjected to feature reduction by K-PCA sequentially followed by the stacking LSTM with XGBoost. The parameters used for K-PCA is listed in the Table I.

TABLE I. EXPERIMENTAL PARAMETERS

Experimental Parameters for K-PCA		Values
n_components		94.3
Kernel		rbf
Gamma		0.04
Alpha		0.03
fit_inverse_transform		True
eigen_solver		Auto
n_jobs		-1

The dataset is split into 80/20 for training and testing. The results obtained by both proposed methods are shown in Table II.

TABLE II. PERFORMANCE OF PROPOSED METHODS

Classifier	Accuracy	Precision	F1-score	Recall	No. of Features
EXG-LSTM	92.1	90.6	94.0	92.7	252
KPCA+EXG-LSTM	94.3	92.0	98.0	94.9	135

The proposed model EXG-LSTM reached an accuracy of 92.1%, a precision of 90.6%, a recall of 92.7%, and an F1-score of 94 % with the features 252. The second method performed exceptionally well when applying K-PCA compared with the result of EXG-LSTM by accomplishing an accuracy of 94.3%, a precision of 92 %, a recall of 94.9%, and with an F1-score of 98% by reducing the feature into 135. The proposed methods' results are compared with the state-of-the-art techniques presented in Table II. Using the cardiologist as the gold standard, the attempt is made to lessen this difference by utilizing ML methods.

TABLE III. PERFORMANCE OF THE PROPOSED METHODS WITH STATE-OF-THE-ART METHODS

Authors	Classifier	Accuracy	Class
Zuo et al. [55] (2008)	KDF-WKNN	70.66	2
Kohli et al. [56] (2007).	OAK	73.40	2
Niazi et al. [57] (2015).	KNN with IFSFS	73.80	2
Prathibhamol et al. [58] (2017).	P-CA-CRA	80.00	16
Pandey et al. [59] (2019).	RNN	83.10	2
Durga [60] (2021).	DT	84.13	-
Jadhav et al. [39] (2014).	MLP	86.67	2
Mitra et al. [61] (2013).	CFS+LM	87.71	2
Raut et al. [62] (2008).	MLPNN	90.00	7
Mustaqeem et al. [44] (2018).	OAD	92.07	16
Khan et al. [45] (2021).	PCA + LSTM	93.50	16
<b>Proposed Method 1</b>	<b>EXG-LSTM</b>	<b>92.10</b>	<b>16</b>
<b>Proposed Method 2</b>	<b>KPCA+EXG-LSTM</b>	<b>94.30</b>	<b>16</b>

Table III clearly reveals the superior performance of our proposed methods. It is worth noting that our proposed methods are capable of classifying 16 classes of arrhythmia with the highest accuracy of 94.30 when compared with state-of-the-art methods as depicted in Fig. 4.

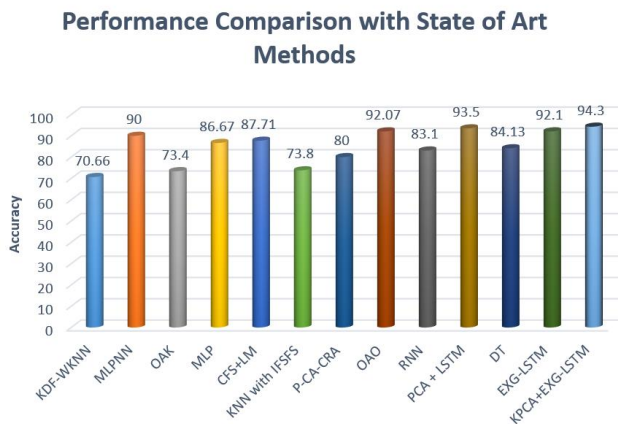


Fig. 4. Proposed models with state-of-the-art methods.

Arrhythmias are potentially fatal. Hence, it is imperative that prediction and analysis be employed in the medical profession with the greatest possible accuracy. The proposed approach highlights these issues by selecting a few notable features using a feature selection mechanism, which enhances classification efficiency.

#### V. CONCLUSION

A novel hybrid DL pipeline for multi-class Arrhythmia classification has been proposed. This approach involves selecting maximally distinct features using the feature selection technique and helps improve the performance of the classification of arrhythmia data. The work implied Kernel-PCA and a Stacked LSTM DL technique with the fusion of XGBoost for classification to determine whether an arrhythmia was present or absent. This research work aims to upgrade the performance of LSTM through the novel pipeline with XGBoost. It was good in its execution and recorded a high accuracy of 92.1%. Concurrently, feature reduction was done, and the performance of the ensemble of LSTM is monitored. Both the proposed models performed well. Among them, EXG-LSTM with K-PCA has reached the highest accuracy of 94.3%. This research works on stacking LSTM shows that the strength of LSTM was boosted by ensembling, and showed better results in classifying the classes of arrhythmia efficiently. Arrhythmias are deadly disorders. Hence prediction and analysis must be exceedingly accurate before being applied in the medical field. The suggested method draws attention to these problems by choosing several standout features with an improved feature selection mechanism, which enhances classification efficiency with reduced number of features of 135. The future direction of the research would be to construct an interpretable ML classifier to augment the classifier's accuracy that will serve as an effective handy diagnostic tool for physicians.

#### REFERENCES

- [1] M. Nishizaki, "Life-threatening arrhythmias leading to syncope in patients with vasospastic angina", *Journal of Arrhythmia*, 2017, 33(6), 553-561.
- [2] R. J. Koene, W.O. Adkisson, & D. G. Benditt, "Syncope and the risk of sudden cardiac death: Evaluation, management, and prevention" *Journal of arrhythmia*, 2017, 33(6), 533-544.
- [3] M. M. Al Rahhal, Y. Bazi and H. A. Hichri. "Deep learning approach for active classification of electrocardiogram signals," *Information Sciences*, 2016, vol. 345, pp. 340-354.
- [4] A. Krizhevsky, I. Sutskever, & G.E. Hinton, "Imagenet classification with deep convolutional neural networks", *Communications of the ACM*, 2017, 60(6), 84-90.
- [5] H. Sugumar, S. Prabhu, A. Voskoboinik, & P.M. Kistler, "Arrhythmia induced cardiomyopathy. *Journal of arrhythmia*, 2018, 34(4), 376-383.
- [6] A. Mustaqeem, S. M. Anwar and M. Majid. "Multiclass classification of cardiac arrhythmia using improved feature selection and SVM invariants," *Computational and Mathematical Methods in Medicine*, vol. 2018, pp. 1-11.
- [7] E. Alickovic and A. Subasi. "Medical decision support system for diagnosis of heart arrhythmia using DWT and random forests classifier," *Journal of Medical Systems*, 2016, vol. 40, no. 108, pp. 1-12.
- [8] P. Rajpurkar, A.Y. Hannun, M. Haghpahani, C. Bourn, & A.Y. Ng, "Cardiologist-level arrhythmia detection with convolutional neural networks". *arXiv preprint arXiv:1707.01836*.
- [9] S. Chen, A. Li, & J.M. Roveda, *Artificial Intelligence in Cardiac Arrhythmia Classification*. learning, 120, 268-275.
- [10] S. U. Kumar and H. H. Inbarani, "Neighborhood rough set-based ECG signal classification for diagnosis of cardiac diseases," *Soft Computing*, 2017, vol. 21, no. 16, pp. 4721-4733.
- [11] Y. Ozbay and B. Karlik "A recognition of ECG arrhythmias using artificial neural networks," in *Proc. IEEE Engineering in Medicine and Biology Society, Istanbul, Turkey*, pp. 1680-1683.
- [12] Z. Ebrahimi, M. Loni, M. Daneshlab, & A. Gharehbaghi, "A review on deep learning methods for ECG arrhythmia classification". *Expert Systems with Applications: 2020, X, 7, 100033*.
- [13] C. G. Nayak, G. Seshikala and U. Desai, "Identification of arrhythmia classes using machine-learning techniques", *International Journal of Biology and Biomedicine*, 2016, vol. 1, pp. 48-53.
- [14] M. Mary Shanthi Rani, P. Chitra, S. Lakshmanan, M. Kalpana Devi, R. Sangeetha, & S. Nithya, "DeepCompNet: A Novel Neural Net Model Compression Architecture", *Computational Intelligence and Neuroscience*, 2022, <https://doi.org/10.1155/2022/2213273>
- [15] N. Karthikeyan, M. S. Rani, "ECG Classification Using Machine Learning Classifiers with Optimal Feature Selection Methods", *In Evolutionary Computing and Mobile Sustainable Networks*, 2022, pp. 277-289. Springer, Singapore. [https://doi.org/10.1007/978-981-16-9605-3\\_19](https://doi.org/10.1007/978-981-16-9605-3_19)
- [16] R. Sangeetha, M. Mary Shanthi Rani, R. Joseph R, "Optimized Deep Neural Network for Tomato Leaf Diseases Identification", *In International Advanced Computing Conference, 2021*, pp. 562-576, Springer, Cham. [https://doi.org/10.1007/978-3-030-95502-1\\_42](https://doi.org/10.1007/978-3-030-95502-1_42)
- [17] S. Nithya, & M.S. Rani, "Deep Learning Model for Arrhythmia Classification with 2D Convolutional Neural Network", *In Innovations in Information and Communication Technologies*, 2022, pp.1-11, Springer, Singapore. [https://doi.org/10.1007/978-981-19-3796-5\\_1](https://doi.org/10.1007/978-981-19-3796-5_1)
- [18] S. Nithya, & M.S.Rani, "Stacked Variational Autoencoder in the Classification of Cardiac Arrhythmia using ECG Signals with 2D-ECG Images", *In International Conference on Intelligent Innovations in Engineering and Technology*, 2022, pp. 222-226). *IEEE Xplore*. <https://doi: 10.1109/ICIET55458.2022.9967575>.
- [19] S. Gupta, R.Sangeeta, R.S. Mishra, G. Singal, T. Badal, & D.Garg, "Corridor segmentation for automatic robot navigation in indoor environment using edge devices", *Computer Networks*, 2020, 178, 107374. <https://doi.org/10.1016/j.comnet.2020.107374>
- [20] S. Nithya, & M.S. Rani, "XACML: Explainable Arrhythmia Classification Model Using Machine Learning", *In International*

- Advanced Computing Conference, 2022, pp. 219-231, Cham: Springer Nature Switzerland.
- [21] C. H. Hsing, "A hybrid intelligent model of analyzing clinical breast cancer data using clustering techniques with feature selection," *Applied Soft Computing*, 2014, vol. 20, pp. 4–14.
- [22] E. R. Hruschka and N. F. Ebecken, "Extracting rules from multilayer perceptron's in classification problems: A clustering-based approach," *Neurocomputing*, 2006, vol. 70, no. 3, pp. 384–397.
- [23] M. Nilashi, "A soft computing approach for diabetes disease classification," *Health Informatics Journal*, 2016, vol. 24, no. 4, pp. 379–393.
- [24] M. Nilashi, O. Ibrahim and A. Ahani, "Accuracy improvement for predicting Parkinson's disease progression," *Scientific Reports*, 2016, vol. 6, no. 1, pp. 34–181.
- [25] M. Nilashi, O. Ibrahim, H. Ahmadi and L. Shahmoradi. "A knowledge-based system for breast cancer classification using the fuzzy logic method," *Telematics and Informatics*, 2017, vol. 4, no. 34, pp. 133–144.
- [26] K. Polat. "Classification of Parkinson's disease using a feature weighting method on the basis of fuzzy C-means clustering," *International Journal of Systems Science*, 2012, vol. 4, no. 4, pp. 597–609.
- [27] A. Ozçift, A. Random forests ensemble classifier trained with data resampling strategy to improve cardiac arrhythmia diagnosis. *Computers in biology and medicine*, 2011, 41(5), 265-271.
- [28] E. Namsrai, T. Munkhdalai, M. Li, J.H. Shin, O.E. Namsrai, K.H.A. & Ryu, "A feature selection-based ensemble method for arrhythmia classification". *Journal of Information Processing Systems*, 2013, 9(1), 31-40.
- [29] H. A. Guvenir, B. Acar, G. Demiroz, & A. Cekin, "A supervised machine learning algorithm for arrhythmia analysis", In *Computers in Cardiology*, 1997, (pp. 433-436). IEEE.
- [30] Y. Freund, R E. Schapire, "A decision-theoretic generalization of online and an application to boosting", *Journal of computer and system sciences*, 55(1), 119-139. (1997). <https://doi.org/10.1006/jcss.1997.1504>
- [31] J H. Friedman, "Greedy function approximation: a gradient boosting machine", *Annals of Statistics*, 2001, 1189-1232.
- [32] R. Khemchandani, S. Chandra S, "Twin support vector machines for pattern classification", *IEEE Transactions on pattern analysis and machine intelligence*, 29(5), 905-910, 2007. <https://doi.org/10.1109/TPAMI.2007.1068>
- [33] S. Dhyani, A. Kumar, & S. Choudhury, "Analysis of ECG-based arrhythmia detection system using machine learning". *MethodsX*, 2023, 10, 102195.
- [34] M. Mohanty, S. Sahoo, P. Biswal, & S. Sabut, "Efficient classification of ventricular arrhythmias using feature selection and C4. 5 classifier", *Biomedical Signal Processing and Control*, 2018, 44, 200-208.
- [35] S. S. Yadav, & S. M. Jadhav, Detection of common risk factors for diagnosis of cardiac arrhythmia using machine learning algorithm. *Expert systems with applications*, 2021, 163, 113807.
- [36] C. U. Kumari, A. S. D. Murthy, B. L. Prasanna, M. P. P. Reddy, & A. K. Panigrahy, "An automated detection of heart arrhythmias using machine learning technique: SVM", *Materials Today: Proceedings*, 2021, 45, 1393-1398.
- [37] F. Taher, H. Alshammari, L. Osman, M. Elhoseny, A. Shehab, & E. Elayat, "Cardiac Arrhythmia Disease Classifier Model Based on a Fuzzy Fusion Approach". *Computers Materials & Continua*, 2023, 75(2), 4485.
- [38] M. Ayar, M., & S. Sabamoniri, "An ECG-based feature selection and heartbeat classification model using a hybrid heuristic algorithm", *Informatics in Medicine Unlocked*, 2018, 13, 167-175.
- [39] S. Jadhav, S. Nalbalwar, & A. Ghatol, "Feature elimination based random subspace ensembles learning for ECG arrhythmia diagnosis", *Soft Computing*, 2014, 18, 579-587.
- [40] M. A. Khan, M.R. Karim, & Y. Kim, "A two-stage big data analytics framework with real world applications using spark machine learning and long short-term memory network", *Symmetry*, 2018, 10(10), 485.
- [41] S. B. Itzhak, S. S. Ricon, S. Biton, J. A. Behar, & J. A. Sobel, Effect of temporal resolution on the detection of cardiac arrhythmias using HRV features and machine learning. *Physiological Measurement*, 2022, 43(4), 045002.
- [42] K. C. Chang, P. H. Hsieh, M. Y. Wu, Y. C. Wang, J. Y. Chen, F. J. Tsai, ... & T.C. Huang, Usefulness of machine learning-based detection and classification of cardiac arrhythmias with 12-lead electrocardiograms. *Canadian Journal of Cardiology*, 2021, 37(1), 94-104.
- [43] Y. Xu, R. Guo, L. Wang, "A twin multi-class classification support vector machine", *Cognitive Computation*, 2013, 5(4), 580-588. <https://doi.org/10.1007/s12559-012-9179-7>
- [44] A. Mustaqeem, S.M. Anwar, M. Majid, "Multiclass classification of cardiac arrhythmia using improved feature selection and SVM invariants", *Computational and mathematical methods in medicine*, 2018. <https://doi.org/10.1155/2018/7310496>
- [45] M. A.Khan, Y. Kim, "Cardiac arrhythmia disease classification using LSTM deep learning approach", *CMC-COMPUTERS MATERIALS & CONTINUA*, 2021, 67(1), 427-443.
- [46] G. Chen, & J.Chen, "A novel wrapper method for feature selection and its applications", *Neurocomputing*, 2015, 159, 219-226.
- [47] G. Chandrashekar, & F. Sahin, "A survey on feature selection methods", *Computers & Electrical Engineering*, 2014, 40(1), 16-28.
- [48] S. Celin, & K. Vasanth, "ECG signal classification using various machine learning techniques", *Journal of medical systems*, 2018, 42(12), 241.
- [49] J. Park, M.Kang, J.Gao, Y.Kim, & K.Kang, "Cascade classification with adaptive feature extraction for arrhythmia detection", *Journal of medical systems*, 2017,41,1-12.
- [50] P. Amorim, T. Moraes, D. Fazanaro, J. Silva, & H. Pedrini, "Shearlet and contourlet transforms for analysis of electrocardiogram signals", *Computer Methods and Programs in Biomedicine*, 2018, 161, 125-132.
- [51] T. F Romdhane, & M.A. Pr, "Electrocardiogram heartbeat classification based on a deep convolutional neural network and focal loss", *Computers in Biology and Medicine*, 2020, 123, 103866.
- [52] M.Sharma, R.S. Tan, & U.R. Acharya, "Automated heartbeat classification and detection of arrhythmia using optimal orthogonal wavelet filters", *Informatics in Medicine Unlocked*, 2019, 16, 100221.
- [53] Y. Chen, X. Wang, Y. Jung, V. Abedi, R. Zand, M. Bikak, & M. Adibuzzaman, "Classification of short single-lead electrocardiograms (ECGs) for atrial fibrillation detection using piecewise linear spline and XGBoost", *Physiological measurement*, 2018, 39(10), 104006.
- [54] H. Assodiky, I. Syarif, & T. Badriyah, "Arrhythmia classification using long short-term memory with adaptive learning rate", *EMITTER International Journal of Engineering Technology*, 20158, 6(1), 75-91.
- [55] W. M. Zuo, W. G. Lu, K. Q. Wang, H. Zhang, "Diagnosis of cardiac arrhythmia using kernel difference weighted KNN classifier". In *Computers in Cardiology* (pp. 253-256). IEEE, 2008. <https://doi.org/10.1109/CIC.2008.4749025>
- [56] N. Kohli, N.K. Verma, A. Roy A, "SVM-based methods for arrhythmia classification in ECG", In *International conference on computer and communication technology (ICCCCT)* (pp. 486-490). IEEE, 2010. <https://doi.org/10.1109/ICCCCT.2010.5640480>
- [57] K. A. K. Niazi, S. A. Khan, A. Shaukat, M. Akhtar M, "Identifying the best feature subset for cardiac arrhythmia classification", In *2015 Science and Information Conference (SAI)* (pp. 494-499). IEEE, 2015. <https://doi.org/10.1109/SAI.2015.7237188>
- [58] Cp P, A. Suresh, G. Suresh, "Prediction of cardiac arrhythmia type using clustering and regression approach (P-CA-CRA)", In *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (pp. 51-54). IEEE, 2017. <https://doi.org/10.1109/ICACCI.2017.8125815>
- [59] S. K. Pandey, R.R. Janghel, "ECG arrhythmia classification using artificial neural networks", In *Proceedings of 2nd International Conference on Communication, Computing and Networking* (pp. 645-652). Springer, Singapore. 2019. [https://doi.org/10.1007/978-981-13-1217-5\\_63](https://doi.org/10.1007/978-981-13-1217-5_63).
- [60] S. Durga, E. Daniel, S.D. Kanmani J.M. Philip, "Cardiac arrhythmia classification using sequential feature selection and decision tree

- classifier method”, International Journal of Innovative Computing and Applications, 2021, 12(4), 175-182.
- [61] M. Mitra, R.K. Samanta, “Cardiac arrhythmia classification using neural networks with selected features”, *Procedia Technology*, 2013, 10, 76-84. <https://doi.org/10.1016/j.protcy.2013.12.33>.
- [62] R. D.Raut, S.V. Dudul, “Arrhythmias classification with MLP neural network and statistical analysis”, In 2008 First International Conference on Emerging Trends in Engineering and Technology (pp. 553-558). IEEE, 2008. <https://doi.org/10.1109/ICETET.2008.260>
- [63] D. Dua, C. Graff, “UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Irvine”, CA: University of California, School of Information and Computer Science. (2019).
- [64] M. Kallas, C. Francis, L. Kanaan, D. Merheb, P. Honeine, H. Amoud, Multi-class SVM classification combined with kernel PCA feature extraction of ECG signals. In 2012 19th International Conference on Telecommunications (ICT) (pp. 1-5). IEEE, 2012. <https://doi.org/10.1109/ICTEL.2012.6221261>
- [65] S. Yang, H. Shen, “Heartbeat classification using discrete wavelet transform and kernel principal component analysis”, In IEEE 2013 Tencon-Spring (pp. 34-38). IEEE, 2013. <https://doi.org/10.1109/TENCONSpring.2013.6584412>
- [66] N. Merrill, C. Olson, “Unsupervised Ensemble-Kernel Principal Component Analysis for Hyperspectral Anomaly Detection”, In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 112-113.
- [67] Y. Dong, Y. Zhang, F. Liu, & X. Cheng, “Reservoir Production Prediction Model Based on a Stacked LSTM Network and Transfer Learning. *ACS omega*, 2021, 6(50), 34700-34711.
- [68] A. Sahar, & D. Han, “An LSTM-based indoor positioning method using Wi-Fi signals”, In Proceedings of the 2nd International Conference on Vision, Image and Signal Processing, 2018, pp. 1-5.
- [69] T. I. Poznyak I. Chairez Oria, A.S. Poznyak, “Background on dynamic neural networks”, *Ozonation and Biodegradation in Environmental Engineering*, 2019, 57-74.
- [70] X. Yang, Z. Chen Z, “A Hybrid Short-Term Load Forecasting Model Based on CatBoost and LSTM”, In 2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP) (pp. 328-332). IEEE, 2021. <https://doi.org/10.1109/ICSP51882.2021.9408768>
- [71] T. Chen, 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. 2016. <https://doi.org/10.1145/2939672.2939785>
- [72] R. Lorbieski, S.M. Nassar, “Impact of an Extra Layer on the Stacking Algorithm for Classification Problems”, *J. Comput. Sci.*, 14(5), 613-622. 2018. DOI: 10.3844/jcssp.2018.613.622
- [73] S. Sang, F. Qu, P. Nie, “Ensembles of Gradient Boosting Recurrent Neural Network for Time Series Data Prediction”, *IEEE Access*. 2021. <https://doi.org/10.1109/ACCESS.2021.3082519>

# A Versatile Shuffle Resource Units Recomputation Algorithm for Uplink OFDMA Random Access

Azyyati Adiah Zazali, Shamala Subramaniam, Zuriati Ahmad Zukarnain, Abdullah Muhammed  
Department of Communication Technology and Network, Faculty of Computer Science and Information Technology,  
Universiti Putra, Malaysia, Serdang, 43400, Selangor Darul Ehsan, Malaysia

**Abstract**—IEEE 802.11ax introduces Uplink Orthogonal Frequency Division Multiple Access (OFDMA)-based Random Access (UORA), a novel feature for facilitating random channel access in Wireless Local Area Networks (WLANs). Similar to the conventional random access scheme in WLANs, UORA employs the OFDMA backoff (OBO) procedure to access the channel's Resource Units (RUs) and selects a random OBO counter within the OFDMA contention window (OCW) range. The Access Point (AP) can determine and communicate this OCW range to each station (STA). Multiple STAs accessing RUs result in transmission failure due to RU collisions, which occur when specific RUs remain unassessed by any STA, leading to wastage. Efforts to optimize channel efficiency require minimizing both collisions and idle RU despite the challenges arising from UORA's distributed and random nature. The Fisher-Yates shuffle algorithm introduces a random uniform distribution strategy for managing RU allocations among STAs. The results demonstrate that this approach enables STAs to access RUs in a distributed manner, effectively reducing idle and wasted RUs, especially in scenarios involving a limited number of STAs. Furthermore, this approach effectively mitigates collisions among STAs, even in scenarios with a more significant number of STAs.

**Keywords**—IEEE 802.11ax; OFDMA; UORA; random access; backoff; resource units allocation; multi-user

## I. INTRODUCTION

Developing MAC (Medium Access Control) protocols that can support large-scale networks with low-power devices elevate the growing number of Internet of Things (IoT) devices; hence, secure communication is essential. OFDMA is a wireless communication technique that allows multiple users to transmit data simultaneously over the same frequency band [1, 2]. Utilizing OFDMA allows sending (e.g., power) to utilize only a fraction of the bandwidth, facilitating simultaneous transmissions, minimizing MAC congestion and reducing overhead, enhancing data transmission efficiency over dense networks, and reducing time wastage.

OFDMA functions in two distinct modes: scheduled access (SA) and random access (RA) [3]. While random access methods like Distributed Coordination Function (DCF) or Enhanced Distributed Channel Access (EDCA) were employed to manage or distribute radio resources in previous WLAN standards, they do not apply to the OFDMA system [1]. UORA is a new feature for random channel access introduced in IEEE 802.11ax.

The channel split into sub-carrier groups known as Resource Units (RUs) in the UORA process. These comprise the minimum OFDMA RUs that STAs need to reach the channel and transmit a frame. With various RUs, multiple STAs can send data packets simultaneously. Each STA chooses a random OBO counter from the OCW value and reduces it by the number of RUs available for UORA to send a particular control frame called a trigger frame (TF). The TF carries information such as the identity information of each STA that may take part in the UL multi-user transmission, the transmission duration, the RUs allocation for each STA, and other helpful information. If the decreased OBO counter drops to zero or less, the STA can send the TF with any available RU. UORA can flexibly control the OCW range based on the number of contending STAs, unlike DCF and EDCA, where the range of contention window (CW) is predetermined.

OCW range is crucial to the UORA performance and primarily depends on the number of contending STAs, but it is challenging for the AP to accurately and quickly estimate or keep track of the number of contending STAs without a specific signaling mechanism.

The Wi-Fi standard random access protocol seldom enables the advantages of reducing network congestion and channel access delay because of the severe frame collision brought by the crowded network conditions to initiate the OFDMA uplink broadcasts. Most studies consider saturation network throughput, but 802.11ax nodes' access delay needs careful examination due to their dependency on AP schedules. In short, the operation of UORA is based on the OCW range and OBO counter values, showing that the random modes of operation make the STAs compete to access the RUs in order to send their UL request during the random selection of one of the 26-RU [4, 5].

The organization of the remaining sections of this article is as follows. Section II delves into the discussion of related open issues concerning UORA. Section III presents the problem formulation. The proposed algorithm is presented in Section IV. The performance evaluation and discussion are in Section V. The conclusion is provided in Section VI.

## II. RELATED OPEN ISSUES IN UORA

Numerous strategies have been put forth in the literature to enhance UORA effectiveness, such as grouping, joint, and clustering [6-12]. By aiming to give high channel efficiency, [6, 7] developed an adaptive grouping scheme on UORA. The study [6] consider a target wake time (TWT) to reduce

transmission collision. At the same time, [7] proposed a Buffer State Report (BSR)-based Two-stage Mechanism-based adaptive STA grouping scheme (BTM) that analyses the relationship between group size and the RU efficiency in an ultra-dense wireless network, mainly when RUs properties are not identical. The proposed spatial clustering group division-based OFDMA (SCGD-OFDMA) by [8] allows the head STA and multiple STAs to compete for the channel resource where the number of STAs is up to 200 when used for random access. The AP in the Grouping-based UORA (G-UORA) method by [9] divides users based on the utility and distributes resources accordingly using the utility prediction model and matching the utility-based resource allocation algorithm. (OFDMA)-based joint reservation and cooperation MAC (OJRC-MAC) protocol proposed in [10] combines channel reservation and cooperation to reduce access collisions and increase transmission dependability, outperforming the basic UORA. Another protocol outperforms the basic UORA by [11] splitting network STAs into spatial groups based on neighbor channel sensing capacity. Designing an information collection system utilizes a probability-based two-level buffer state report, where intra-group STAs transmit data with minimal power consumption. Simultaneously, the AP strategically organizes two spatial groups to prevent interference. The modification by [12] from the 2019 IEEE 802.11be Wi-Fi standard combined with a novel resource allocation algorithm provides effective real-time communications for the uplink OFDMA feature.

UORA introduces the OBO counter for the operation of multi-user transmission. An important observation is that the backoff procedure for UORA differs from that of DCF or EDCA. While DCF and EDCA perform backoff in the time domain to decide when to transmit, UORA's backoff procedure is two-dimensional, which simultaneously determines both the RUs to occupy in the frequency domain and the transmission time to produce high-efficiency results [1, 13, 14, 15, 16]. The research in [1] demonstrates a noteworthy increase in throughput through a straightforward modification to the OBO counter. However, implementing this suggested mechanism in real-world WLANs with practical settings becomes impractical due to the complexities associated with the OBO control rule. The CRUI (Collision Reduction and Utilization Improvement) method, introduced in [13], improves UORA performance by reducing transmission collisions. This enhancement involves increasing backoff times and utilizing opportunistic sub-channel hopping. Notably, the CRUI scheme ensures that it does not degrade the performance of UL transmissions and prioritizes distributed real-time STA transmissions. Retransmission number aware channel access (RNACA) increases throughput when there are more random access RUs than STAs and lowers packet latency [14]. By optimizing parameters like  $CW_{min}$  and  $CW_{max}$ , RNACA significantly increases the throughput of the maximum number of transmission trials (MNTTSTAs can conduct complementary transmission without backoff using the probability complementary transmission scheme (PCTS) described in [15]. However, it is essential to note that this scheme is applicable only in WLANs with lower mobile station numbers than random access RUs, as STAs must choose RUs based on OFDMA-based backoff counters, causing retransmission delays. In [16], developing a new uplink hybrid UORA (H-

UORA) OFDMA access mechanism introduces an RU-sensing slot, enabling additional channel sensing to minimize transmission collisions further. However, H-UORA needs a much finer carrier sensing circuit for the current 802.11ax amendment.

The 802.11ax network throughput can be optimized through RU allocation strategies, as demonstrated in [17, 18, 19, 20, 21]. The research in [17] examined the effects of various RA RU and SA RU distributions on the MAC layer performance, while [18] created a new RU distribution scheduler for managing access that features a closed-loop feedback controller with proportional gain. In [19], the research algorithm achieves a delay of less than one millisecond with a remarkably high level of reliability. On the other hand, in [20], OFDMA transmissions ensure reliable and dependable communication, especially in the presence of interference. Both studies demonstrate the ability to support real-time applications. The study in [21] proposed a channel access scheme for next-generation vehicle-to-everything (V2X) systems that expands backward compatibility with IEEE 802.11-based extension.

Fair Allocation and Effective Utilisation of RUs (FAEU-RUs) protocol in [5] handles the OFDMA MU communications based on the two factors. First is the fairness criterion of ensuring that all STAs in UL reasonably access the RUs, and second is the practical criterion of ensuring that the RUs are optimally allocated and used. The suggested solution in [3] guarantees fairness in RU access and requires minimal overhead, producing results close to optimal. However, it does not effectively adapt to traffic needs.

UORA and RU allocation are a big area to discover because of the drawback of this access mode, which is the high rate of collisions due to competition. Researchers continuously refine advancements in RU allocation for OFDMA as indicated by ongoing research on related open issues. This research focuses on the RU allocation for UORA.

### III. PROBLEM FORMULATION

The IEEE 802.11ax standard [22] defines scheduled and random access as two distinct kinds of uplink multi-user (MU) OFDMA operations. In scheduled access, each STA uses BSR signaling to ask the AP for authorization to transmit while the STAs share the OFDMA RUs without causing any contention. The AP then transmits a TF carrying the scheduling data to allot a dedicated RU to a particular STA. On the other hand, in a random access mode, the STA acquires the RU by the UORA mechanism in a contention-based way.

Fig. 1 illustrates a multi-user wireless communication network model with a single AP and a plurality of STAs based on [4] invention. Table I presents device attributes, which are in Fig. 1. The AP represents the access points, while STA1, STA2, STA3, STA4, STA5, STA6, and STA7 represent various wireless communication devices or STAs.

The UORA method operated by the AP shows that every  $n^{\text{th}}$  TF for random access transmitted by the AP includes at least one RU for random access available to 20 MHz operating STAs, where N is a positive integer. In other words, every Nth TF for random access contains at least one RU for random



access in the primary 20 MHz channel and is unrestricted from being used for 20 MHz operating STAs. A 20 MHz operating STA is allowed to reach the AP with the UORA mechanism when receiving TF for random access.

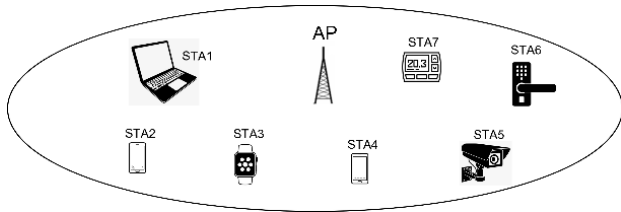


Fig. 1. Network model of a multi-user wireless communication network.

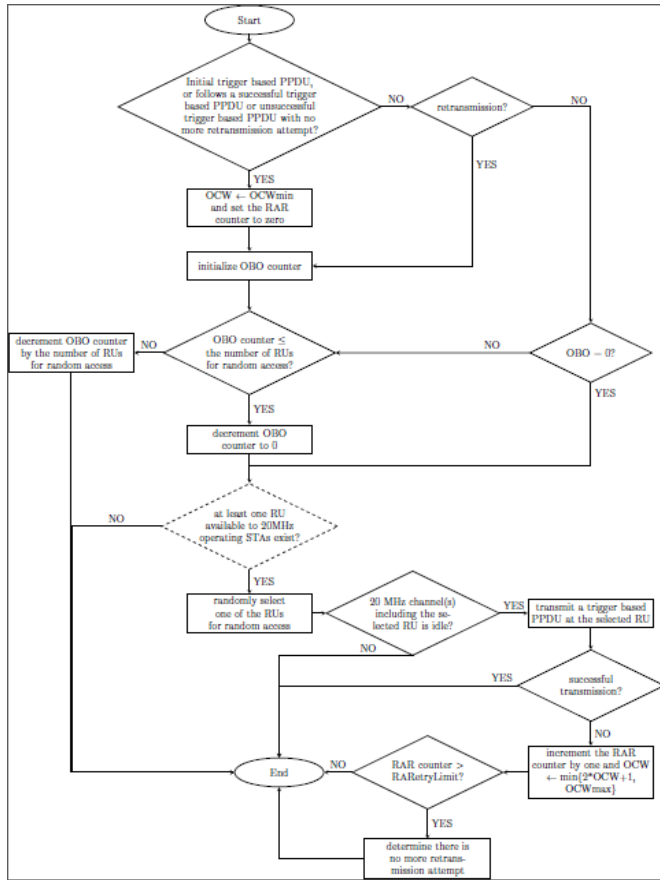


Fig. 2. UORA method operated by a 20MHz station.

The flow chart in Fig. 2 shows an example UORA method operated by a 20 MHz operating station with a detailed explanation in Table II. We assume this scenario occurs between the AP and any available STA capable of operating with a 20 MHz channel width, particularly STA3, STA6, and STA7, as indicated in Table I, since these STAs can only operate using this channel width.

TABLE I. DEVICES ATTRIBUTES

	STA1	STA5	STA2	STA4	STA3	STA6	STA7
Device type	laptop	CCTV	smartphone	smartwatch	Door lock	Thermostat	
QoS requirement	high			low			
Power	Low power	Concern about		Compassionate power			

management	saving	power consumption	consumption
Channel width (MHz)	20, 40, 80, 80+80, 160	20, 40, 80	20

TABLE II. DETAILING ON STEPS ACCORDING TO THE FLOW CHART IN FIG. 2

Step	Event
1	20 MHz operating STA receives a TF for random access from the AP.
2	20 MHz operating STA determines its UL transmission as an initial trigger-based PPDU transmission, follows a successful trigger-based PPDU transmission, or follows an unsuccessful trigger-based PPDU transmission for which there is no more retransmission attempt.
2.1	If 20 MHz operating STA UL transmission fulfils Step 2, the STA sets OCW value to OCWmin value and the Random Access Retry (RAR) counter to zero.
2.2	If 20 MHz operating STA UL transmission is unfit, Step 2, 20 MHz operating STA, continues to check if its UL transmission is a retransmission of an unsuccessful trigger-based PPDU transmission.
2.2.1	If 20 MHz operating STA UL transmission is retransmission, the UORA method proceeds to Step 3
2.2.2	If 20 MHz operating STA UL transmission is not retransmission, 20 MHz operating STA determines if the OBO counter is equal to value zero.
2.2.2.1	If the 20 MHz operating STA OBO counter equals zero, the UORA method proceeds to Step 5, implying that the 20 MHz operating STA won the contention but did not transmit a trigger-based PPDU in the previously selected RU in Step 1 since one or more 20 MHz channels containing the previously selected RU are considered busy.
2.2.2.2	If the 20 MHz operating STA OBO counter is not equal to zero, the UORA method proceeds to Step 4. This situation implies that the 20 MHz operating STA did not win the contention to access the RUs for random access in the previous TF in Step 1.
3	20 MHz operating STA initializes its OBO counter to a random value in the range of zero and OCW.
4	20 MHz operating STA checks that its OBO counter is smaller or equal to the number of RUs for random access in the previous TF in Step 1.
4.1	If the OBO counter is smaller or equal, 20 MHz operating STA decreases its OBO counter to zero. This situation implies that 20 MHz operating STA wins the random access contention, which proceeds to Step 5.
4.2	If the OBO counter is higher than 0, 20 MHz operating STA decrement its OBO counter by the number of RUs for random access in the received TF in Step 1, which proceeds in Step 11.
5	20 MHz operating STA determines if at least one RU for random access available to 20 MHz operating STA exists in the received TF.
5.1	If 20 MHz operating STA fulfils Step 5, the UORA method proceeds to Step 6.
5.2	If 20 MHz operating STA is unfit in Step 5, proceed to Step 11.
6	20 MHz operating STA randomly selects one of the RUs for random access in the previous TF in Step 1.
7	20 MHz operating STA checks if each of one or more 20 MHz channels, including the selected RU, is idle due to physical and virtual carrier sensing.
7.1	If the selected RU is idle, 20 MHz operating STA transmits a trigger-based PPDU at the selected RU.
7.2	If the selected RU is not idle, proceed with Step 11.
8	The STA operating at 20 MHz determines the successful transmission of the trigger-based PPDU on the selected RU.
8.1	In the event of a successful transmission, proceed to Step 11.
8.2	If an immediate response is not received as solicited by Step 8, consider the transmission unsuccessful and proceed to Step 9.
9	20 MHz operating STA increments the RAR counter by one and sets the OCW value to the minimum of a sum of double the current OCW value plus one and a value of OCWmax.
10	20 MHz operating STA determines if the RAR counter is more significant than a RARRetryLimit threshold, indicating the maximum

	number of random access retransmission attempts.
10.1	If the RAR counter is lesser than RARetryLimit, proceed to Step 11.
10.2	If the RAR counter is enormous than RARetryLimit, 20 MHz operating STA determines there is no more retransmission attempt and then proceeds to Step 11
11	UORA method comes to an end or stops.

TF Info	STA 1 (unassociated)	STA 2 (associated)	STA 3 (associated)	STA 4 (associated)	STA 5 (associated)	STA 6 (associated)	STA 7 (associated)
Random initial OBO value	0	7	7	5	6	5	3
OBO counter	0 - 1 = -1	7 - 8 = -1	7 - 8 = -1	5 - 8 = -3	6 - 8 = -2	5 - 8 = -3	3 - 8 = -5
RU1 (0)	-	-	-	-	-	-	-
RU2 (0)	-	-	Access	-	-	Access	-
RU3 (0)	-	-	-	-	-	-	-
RU4 (0)	-	-	-	Access	-	-	-
RU5 (0)	-	Access	-	-	-	-	-
RU6 (0)	-	-	-	-	-	-	Access
RU7 (0)	-	-	-	-	Access	-	-
RU8 (0)	-	-	-	-	-	-	-
RU9 (2045)	Access	-	-	-	-	-	-

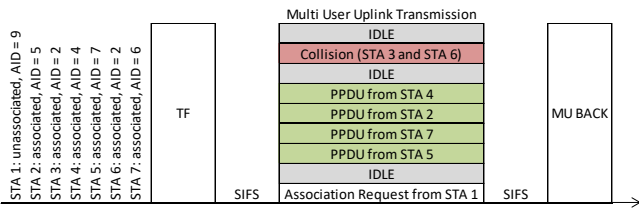


Fig. 3. An example of UORA operation in time and the OBO decrementation after TF reception based on [23].

The usage of the UORA mechanism in a contention-based manner, the STA acquires the RU, as shown in Fig. 3. The AP sends a TF to start the UORA process in the UORA mechanism. The TF includes various data bits, including the associated association identifiers and the eligible random access RUs (RA-RUs) (AIDs). The associated STAs can utilize RA-RUs with the AID number 0, while unassociated STAs can occupy RA-RUs with the AID number 2045 following the 802.11ax standard. The AP may assign some RUs to planned access and others to random access.

However, to concentrate on the efficiency of UORA, the assumption was made that all RUs are qualified for random access without taking scheduled access into account. To be more precise, one RU is assigned to have an AID of 2045 and the rest to have an AID of 0. According to the IEEE 802.11ax standard, the overall number of RUs depends on the channel bandwidth and the number of sub-carriers used for each RU.

Fig. 3 represents an example of the UORA operation. STA 1 is unassociated, and STAs 2-7 are associated. The channel bandwidth is 20 MHz, and 9 RA-RUs consist of eight RUs with AID 0 and one with AID 2045. On receiving the TF, STA 1 decreases its OBO counter by 1, and STAs 2-7 decrease the OBO counter by 8. In this example, the OBO counter for STAs 2-7 becomes ≤ 0, which means the OBO counter is not greater than the number of eligible RUs, and then the STAs sets its OBO counter to 0. Each time OBO = 0, the STAs select a random RA-RU among RU 1-8 to transmit a frame. If there are cases between the STAs where the OBO counter is > 0, the STA cannot access the channel and decreases its OBO counter, waiting for the next contention round upon receiving the next TF. This example also shows that RU can collide and remain idle. STA 3 and STA 6 access the same RU 2, so their transmission can fail due to collisions. Instead, STAs do not

access specific RUs (1, 3, and 8), causing these RUs to become wasted. In case of an unsuccessful transmission, the STAs follow the retransmission procedure shown in Algorithm 1 as in the standard UORA. To ensure channel efficiency, one must minimize the number of collisions and wasted RUs. However, achieving this goal is challenging, given the distributed and random nature of UORA. The backoff procedure in UORA is two-dimensional because it determines which RU to occupy in the frequency domain and simultaneously establishes the transmission time, unlike DCF or EDCA, which performs the backoff procedure in the time domain to determine when to transmit.

### Algorithm 1 Standard UORA

```

OCWmin = 7
OCWmax = 31
if first transmission, then
    OCW = OCWmin;
    OBO = random integer(0, OCW);
else if retransmission, then
    OCW = 2 × OCWold + 1;
    if OCW ≥ OCWmax then
        OCW = OCWmax;
    end if
    OBO = random integer(0, OCW);
end if
Station decrements OBO by number of RU and selects a random
RU for transmission if OBO = 0
    
```

More examples are provided by [4] for different types of STA channel widths to perform the frequency scheduling for OFDMA multi-user transmission in 802.11ax. Frequency scheduling is generally performed based on RU that comprises a plurality of consecutive subcarriers. According to frequency scheduling, a radio communication AP adaptively assigns RUs to a plurality of STA based on the reception qualities of frequency bands of the STAs. The situation makes it possible to obtain a maximum multi-user diversity effect and efficiently perform communication.

While the techniques outlined in this disclosure apply to various wireless communication systems, it is essential to note that, for illustrative purposes, the subsequent descriptions in this disclosure pertain to an IEEE 802.11 WLAN system and its associated terminologies. However, this choice of example should not restrict the scope of this disclosure concerning alternative wireless communication systems. In IEEE 802.11-based WLANs, most networks operate in infrastructure mode, i.e., all or most of the traffic in the network must go through the AP. As such, any STA wishing to join the WLAN must first negotiate the network membership with the AP through association and authentication.

### IV. PROPOSED VERSATILE SHUFFLE RECOMPUTATION RU ALGORITHM

We proposed to change the operation of standard UORA behaviors as in Algorithm 1 to improve its efficiency. To ensure the adaptability of STAs allocation based on the

available RUs, we introduce the versatile shuffle recomputation RU (VSR-RUs) algorithm.

In VSR-RUs, The Fisher-Yates shuffle algorithm [24] shown in Algorithm 2, also known as the Knuth shuffle or the Durstenfeld shuffle, was applied for the RUs recomputation. The Fisher-Yates shuffle algorithm is employed to efficiently generate a random permutation of an array of elements by randomly shuffling them. This shuffle algorithm iterates the array in reverse order and swaps the current element at each step with a randomly chosen element that comes before it (including itself). This process ensures that every element in the original array has an equal chance of being placed in any position within the shuffled array, which means that when applied to UORA, it is the RUs allocation. Introducing randomness at each process step establishes a uniform likelihood of interchanging with any previous element. The algorithm has a time complexity of  $O(n)$ , where  $n$  is the number of elements in the array. It is considered a highly efficient and unbiased method for shuffling arrays.

**Algorithm 2** The Fisher-Yates shuffle algorithm

```

function fisherYatesShuffle(array)
    n = length(array);

    for i from n - 1 down to 1:
        j = random integer(0, i);
        swap(array[i], array[j]);
        OBO = random integer(0, OCW);

function randomInteger(min, max):
    return random(min, max);

function swap(a, b):
    temp = a;
    a = b;
    b = temp;
    
```

Table III shows the STAs RA-RUs allocation in standard OURA in Fig. 3 after applying the Fisher-Yates shuffle algorithm. Recomputing the allocation of RUs prevents collisions within RU 2 while distributing the available RUs among STAs ensures that each STA is guaranteed its allocation without sharing. However, RU 3 and RU 4 remain idle as the number of STAs is < the number of RUs. If an STA OBO counter is greater than 0 and cannot access the channel, it could be due to the STA not being allocated in any available RUs. In such cases, the STA must wait for the next contention round upon receiving the next TF.

Fig. 4 is an example of the Fisher-Yates shuffle mechanism using the RUs array [1, 2, 3, 4, 5, 6, 7, 8, 9]. The mechanism is performed in the following order;

- 1) Randomly select a number  $k$  from 1 to 9, and then swap the  $k^{\text{th}}$  and 9<sup>th</sup> STA. So, if the random number is 4, swap the 4<sup>th</sup> and 9<sup>th</sup> STA in the list.
- 2) Select the following random number from 1 to 8 and swap the chosen STA with the 8th STA. If it is 6, for example, swap the 6<sup>th</sup> and 8<sup>th</sup> STA.

- 3) If the array contains multiple STAs, a random selection will determine which STAs are placed on the list.
- 4) The iteration continues until the permutation is completed, as illustrated in Fig. 4.

After shuffling, the last row of the shuffle displays the output on RUs. This process ensures that the original order of the array is completely randomized while maintaining a uniform distribution of possible outcomes.

TABLE III. THE STAs RUS SELECTION ON STANDARD UORA BEFORE AND AFTER APPLYING ALGORITHM 2

RU #	STAs allocation on RU #	
	Before	After
RU 1	[]	[Associated STA 5 (AID: 0)]
RU 2	[Associated STA 3 (AID: 0), Associated STA 6 (AID: 0)]	[Associated STA 3 (AID: 0)]
RU 3	[]	[]
RU 4	[Associated STA 4 (AID: 0)]	[]
RU 5	[Associated STA 2 (AID: 0)]	[Unassociated STA 1 (AID: 2045)]
RU 6	[Associated STA 7 (AID: 0)]	[Associated STA 6 (AID: 0)]
RU 7	[Associated STA 5 (AID: 0)]	[Associated STA 2 (AID: 0)]
RU 8	[]	[Associated STA 7 (AID: 0)]
RU 9	[Unassociated STA 1 (AID: 2045)]	[Associated STA 4 (AID: 0)]



Fig. 4. Recomputation of array for RUs allocation by using Fisher-Yates shuffle.

V. PERFORMANCE EVALUATION AND DISCUSSION

This section concentrates on the implementation and evaluation of the performance of the proposed VSR-RUs algorithm. A comparison is made against the standard UORA [22]. These evaluations are used to assess the effectiveness of the VSR-RUs algorithm. The Java programming language develops the DES simulation using Eclipse IDE for Java Developers with Indigo Service Release 2 software.

**Algorithm 3** Standard OURA pseudocode structure

```

for a := 1 to NumberofExperiment do
  initialization()
  for i := 1 to maxNoOfStation do
    event[i][j] = time assign; //rand()%(OCW-1)
  end for
  while simulation run do
    if (simulationTime < 60) then
      scheduler()
      update simulationClock()
      if eventType = 1 then
        intendtoTransmit()
      else if eventType = 2 then
        packetGeneration()
      else if eventType = 3 then
        triggerFrame()
      else if eventType = 4 then
        acknowledgment()
      end if
    end if
  end while
  result()
end for

```

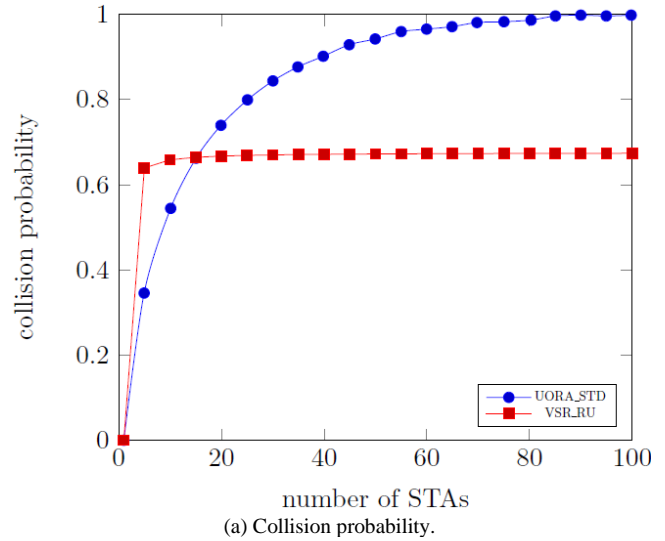
Algorithm 3 shows the pseudocode structure for the novel UORA. The pseudocode begins with the initialization of all the parameters involved. Table IV displays the parameters. Each STA has an event defined, and Table V illustrates these defined events. The Random() function from the mathematical Java library generates the random number, adhering to a uniform distribution. The larger the number generated, the smaller the access probability. During the simulation, as long as the simulation time remains under a minute, the scheduler function will execute, leading to an update of the simulation clock. Subsequently, the event will be chosen based on its type, and the simulation present the outcome afterwards.

TABLE IV. SIMULATION PARAMETERS

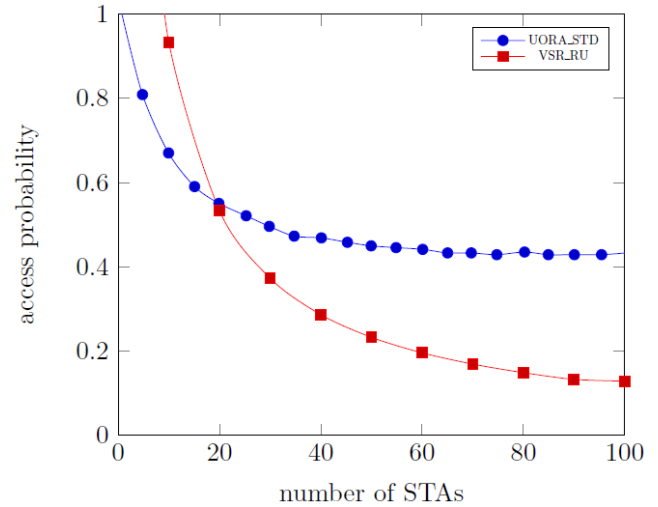
Parameter	Value
Simulation time	60 s
Channel bandwidth	20 MHz
Number of subcarriers RU	26
Number of RU AID = 0	8
Number of RU AID = 2045	1
Number of contending STAs	1~100
Data rate per RU	6.67 Mb/s
Trigger frame length	100µs

TABLE V. UPLINK OFDMA EVENT DEFINE

No	Event	STA	Time Assign
1	Frame transmission intention	1	ev[1][1]
2	Generate random number	1	ev[1][2] = rand();
3	Trigger frame	1	ev[1][3] = transmit TF
4	Acknowledgment	1	ev[1][4] = transmit ACK
No	Event	AP	Time Assign
1	Frame transmission intention	1	evAP[1]
2	Trigger frame	1	evAP[2] = transmit TF
3	Acknowledgment	1	evAP[3] = transmit ACK



(a) Collision probability.



(b) Channel access probability.

Fig. 5. Performance comparison of collision probability and channel access probability between standard UORA and the proposed VSR\_RU.

Fig. 5 compares the collision probability and channel access probability between the standard UORA and VSR\_RU proposed algorithm. Fig. 5(a) shows that the number of collision probabilities for VSR-RUs is at steady state once the number of STAs approaches ten and increases. As the number of STAs increases, the RUs allocation is more uniformly distributed, thus contributing to the decrease in collision.

When examining Fig. 5(b), it becomes evident that with fewer STAs (1-20), the increased access probability of VSR\_RUs has reduced the probability of having idle RUs. This reduction in idle RUs has the potential to enhance overall throughput. Conversely, employing lower access probabilities from VSR\_RU could be a strategic choice for distributing resources among STAs or processes demanding a higher quality of service or performance. This approach would prioritize critical tasks appropriately.

The proposed VSR\_RU algorithm improved the random distribution for allocating STAs to RUs, resulting in lower collisions among STAs. This strategy enables STAs to access RUs in a distributed fashion, which may cause fewer idle and wasted RUs, especially in scenarios with limited STAs. Besides, STA collisions are less when the number of STAs is higher.

## VI. CONCLUSION

The VSR\_RU algorithm, as proposed, employs a random uniform distribution approach using the Fisher-Yates shuffle algorithm for each STA to manage allocations in available RUs. This algorithm operates without requiring any prior information about the number of STAs. The key innovation lies in introducing a shuffled random distribution for assigning STAs to RUs, effectively minimizing collisions among STAs. This approach enables STAs to access RUs in a distributed manner, thereby reducing the occurrence of idle and wasted RUs, particularly in scenarios with a small number of STAs. Moreover, it also mitigates STA collisions in scenarios with many STAs. The adaptability incorporated into the proposed algorithm improves the standard IEEE 802.11ax UORA mechanism.

Further development could involve refining the OBO control rule through a more efficient UORA mechanism within the proposed algorithm. The limitation of the research is that shuffling can take a significant amount of time for massive arrays by applying the Fisher-Yates algorithms. Additionally, reinforcement learning technology can enhance UORA performance in real-world WLAN environments.

## ACKNOWLEDGMENT

This paper and research has been extended support from the Universiti Putra Malaysia Contract Research Grant (Vot No: 6300375).

## REFERENCES

- [1] Y. Kim, L. Kwon, and E.-C. Park, "OFDMA Backoff Control Scheme for Improving Channel Efficiency in the Dynamic Network Environment of IEEE 802.11ax WLANs," *Sensors*, vol. 21, no. 15, 2021, doi: 10.3390/s21151111.
- [2] C. Christopher, "Orthogonal Frequency Division Multiple Access," *In An Introduction to LTE LTE, LTE-Advanced, SAE, VoLTE and 4G Mobile Communications*, pp. 67-85. John Wiley & Sons, Ltd, 2014, doi: 10.1002/9781118818046.ch4.
- [3] K. Kosek-Szott, and K. Domino, "An Efficient Backoff Procedure for IEEE 802.11ax Uplink OFDMA-Based Random Access," *IEEE Access*, vol. 10, pp. 8855-8863, 2022, doi: 10.1109/ACCESS.2022.3140560.
- [4] L. Huang, R. Chitrakar, Y. Urabe, I. Yoshii, "Orthogonal Frequency-Division Multiple Access Communication Apparatus and Communication Method," U.S. Patent 20210282184A1, 9 Sep. 2021.
- [5] S. Brahmi, and M. Yazid, "Towards a Fair Allocation and Effective Utilization of Resource Units in Multi-User WLANs-based OFDMA technology," *Computer Networks*, vol. 224, pp. 109639, 2023, doi: 10.1016/j.comnet.2023.109639.
- [6] J. Bai, H. Fang, J. Suh, O. Aboul-Magd, E. Au, and X. Wang, "Adaptive Uplink OFDMA Random Access Grouping Scheme for Ultra-Dense Networks in IEEE 802.11ax," 2018 IEEE/CIC International Conference on Communications in China (ICCC), Beijing, China, 2018, pp. 34-39, doi: 10.1109/ICCCChina.2018.8641202.
- [7] J. Bai, H. Fang, J. Suh, O. Aboul-Magd, E. Au, and X. Wang, "An adaptive grouping scheme in ultra-dense IEEE 802.11ax network using buffer state report based two-stage mechanism," *China Communications*, vol. 16, no. 9, pp. 31-44, Sept. 2019, doi: 10.23919/JCC.2019.09.003.
- [8] Y. Li, B. Li, M. Yang, and Z. Ya, "A spatial clustering group division-based OFDMA access protocol for the next generation WLAN," *Wireless Networks*, vol. 25, pp. 5083-5097, 2019, doi: 10.1007/s11276-019-02115-2.
- [9] A. Yang, B. Li, M. Yang, Z. Yan, and Y. Xie, "Utility optimization of grouping-based uplink OFDMA random access for the next generation WLANs," *Wireless Networks*, vol. 27, pp. 809-823, 2021, doi: 10.1007/s11276-020-02489-8.
- [10] Y. Zhang, B. Li, M. Yang, Z. Yan, and X. Zuo, "An OFDMA-based joint reservation and cooperation MAC protocol for the next generation WLAN," *Wireless Networks*, vol. 25, pp. 471-485, 2019, doi: 10.1007/s11276-017-1567-1.
- [11] M. Peng, B. Li, Z. Yan, and M. Yang, "A Spatial Group-Based Multi-User Full-Duplex OFDMA MAC Protocol for the Next-Generation WLAN," *Sensors*, vol. 20, no. 14, 2020, doi: 10.3390/s20143826.
- [12] E. Avdotin, D. Bankov, E. Khorov, and A. Lyakhov, "Enabling Massive Real-Time Applications in IEEE 802.11be Networks," 2019 IEEE 30th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Istanbul, Turkey, 2019, pp. 1-6, doi: 10.1109/PIMRC.2019.8904271.
- [13] J. Kim, H. Lee, and S. Bahk, "CRUI: Collision Reduction and Utilisation Improvement in OFDMA-Based 802.11ax Networks," 2019 IEEE Global Communications Conference (GLOBECOM), Waikoloa, HI, USA, 2019, pp. 1-6, doi: 10.1109/GLOBECOM38437.2019.9013337.
- [14] Y. Zheng, J. Wang, J., Q. Chen, and Y. Zhu, "Retransmission Number Aware Channel Access Scheme for IEEE 802.11ax Based WLAN," *Chinese Journal of Electronics*, vol. 29, no. 2, pp 351-360, 2020, doi: 10.1049/cje.2020.01.014.
- [15] J. Wang, M. Wu, Q. Chen, Y. Zheng, and Y. -h. Zhu, "Probability Complementary Transmission Scheme for Uplink OFDMA-based Random Access in 802.11ax WLAN," 2019 IEEE Wireless Communications and Networking Conference (WCNC), Marrakesh, Morocco, 2019, pp. 1-7, doi: 10.1109/WCNC.2019.8885789.
- [16] L. Lanante, C. Ghosh, and S. Roy, "Hybrid OFDMA Random Access With Resource Unit Sensing for Next-Gen 802.11ax WLANs," *IEEE Transactions on Mobile Computing*, vol. 20, no. 12, pp. 3338-3350, 1 Dec. 2021, doi: 10.1109/TMC.2020.3000503.
- [17] S. Bhattarai, G. Naik, and J. -M. J. Park, "Uplink Resource Allocation in IEEE 802.11ax," *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, Shanghai, China, 2019, pp. 1-6, doi: 10.1109/ICC.2019.8761594.
- [18] D. G. Filoso, R. Kubo, K. Hara, S. Tamaki, K. Minami, and K. Tsuji, "Proportional-based Resource Allocation Control with QoS Adaptation for IEEE 802.11ax," *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, Dublin, Ireland, 2020, pp. 1-6, doi: 10.1109/ICC40277.2020.9149111.
- [19] E. Avdotin, D. Bankov, E. Khorov, and A. Lyakhov, "OFDMA Resource Allocation for Real-Time Applications in IEEE 802.11ax Networks," 2019 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom), Sochi, Russia, 2019, pp. 1-3, doi: 10.1109/BlackSeaCom.2019.8812774.
- [20] E. Avdotin, D. Bankov, E. Khorov, and A. Lyakhov, "Resource Allocation Strategies for Real-Time Applications in Wi-Fi 7," 2020 IEEE International Black Sea Conference on Communications and

- Networking (BlackSeaCom), Odessa, Ukraine, 2020, pp. 1-6, doi: 10.1109/BlackSeaCom48709.2020.9234994.
- [21] J. Ahn, Y. Y. Kim, R. Y. Kim, "A Novel WLAN Vehicle-To-Anything (V2X) Channel Access Scheme for IEEE 802.11p-Based Next-Generation Connected Car Networks," *Applied Sciences*, vol. 8, no. 11, 2018, doi: 10.3390/app8112112.
- [22] "IEEE Draft Standard for Information Technology -- Telecommunications and Information Exchange Between Systems Local and Metropolitan Area Networks -- Specific Requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment Enhancements for High Efficiency WLAN," IEEE P802.11ax/D4.0, February 2019, vol., no., pp.1-746, 12 Mar. 2019.
- [23] "IEEE Standard for Information Technology--Telecommunications and Information Exchange between Systems Local and Metropolitan Area Networks--Specific Requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 1: Enhancements for High-Efficiency WLAN," IEEE Std 802.11ax-2021 (Amendment to IEEE Std 802.11-2020), vol., no., pp.1-767, 19 May 2021, doi: 10.1109/IEEESTD.2021.9442429.
- [24] E. Manuel, "Fisher-Yates shuffle," *Archive of Formal Proof*, Nov. 2016.

# The Promise of Self-Supervised Learning for Dental Caries

Tran Quang Vinh, Haewon Byeon\*

Department of Digital Anti-Aging Healthcare (BK21), Inje University, Gimhae 50834, South Korea

**Abstract**—Self-supervised learning (SSL) is a type of machine learning that does not require labeled data. Instead, SSL algorithms learn from unlabeled data by predicting the order of image patches, predicting the missing pixels in an image, or predicting the rotation of an image. SSL has been shown to be effective for a variety of tasks, including image classification, object detection, and segmentation. Dental image processing is a rapidly growing field with a wide range of applications, such as caries detection, periodontal disease progression prediction, and oral cancer detection. However, the manual annotation of dental images is time-consuming and expensive, which limits the development of dental image processing algorithms. In recent years, there has been growing interest in using SSL for dental image processing. SSL algorithms have the potential to overcome the challenges of manual annotation and to improve the accuracy of dental image analysis. This paper conducts a comparative examination between studies that have used SSL for dental caries processing and others that use machine learning methods. We also discuss the challenges and opportunities for using SSL in dental image processing. We conclude that SSL is a promising approach for dental image processing. SSL has the potential to improve the accuracy and efficiency of dental image analysis, and it can be used to overcome the challenges of manual annotation. We believe that SSL will play an increasingly important role in dental image processing in the years to come.

**Keywords**—Machine learning; dental imaging; dental caries; oral diseases

## I. INTRODUCTION

Artificial intelligence (AI) represents a domain within computer science that focuses on crafting intelligent entities, these being systems endowed with the capacities of logical deduction, knowledge acquisition, and independent decision-making. The field of AI has witnessed remarkable achievements in formulating potent methodologies to address a diverse spectrum of challenges, spanning from strategic game playing to intricate medical diagnostics [1]. Interest in the medical application of AI has lately surged due to the impact of this technology on the outcome and caliber of clinical practice during and after the 1980s [2]. Precision medicine, population health, and natural language processing are just a few of the areas of healthcare and medical practice where AI has been researched [3].

Machine learning (ML), a type of AI that allows computers to learn without being programmed. ML algorithms are trained on large datasets of data, and they can then be used to make predictions or decisions. ML is being used in a variety of medical applications, including diagnosis, treatment planning, drug discovery, personalized medicine, and

healthcare management. ML algorithms have been shown to be effective in a variety of tasks, such as detecting cancer, planning radiation therapy, and identifying potential new treatments for diseases [4].

Annotation in medical imaging is the process of labeling or describing medical images with relevant information. This information can be used to train machine learning algorithms to diagnose diseases, plan treatments, or conduct research. There are a variety of ways to annotate medical images, including: Manual annotation, Semi-automated annotation, and Automated annotation [5]. Nonetheless, the substantial expenses tied to acquiring essential specialized annotations often impede endeavors to employ machine learning algorithms for aiding clinical applications. Even partially automated software tools might fall short of significantly alleviating the financial burden associated with annotations. [5]. Self-supervised representation learning [6] is a type of machine learning that learns to represent data without being explicitly labeled. This is in contrast to supervised learning, where the data is labeled with the desired output. Recently, interest in these techniques has increased [7, 8]. In particular, self-supervised representation learning may enhance label effectiveness and performance in scenarios involving the classification of dental caries. In the area of dental caries, this article reviews self-supervised algorithm and compare it to other learning techniques.

## II. MATERIALS AND METHODS

### A. Artificial Intelligence and Machine Learning in Dental Caries

Dental caries, commonly known as tooth decay, is a disease that causes tooth decay by bacteria in the mouth producing lactic acids, which directly harm the tooth surface layer known as the enamel layer. This can progressively lead to a small hole or cavity in the teeth; if not treated, this can cause discomfort, infection, and finally tooth loss [9]. Early detection of carious lesions is necessary for the management of dental caries.

The discipline of dentistry saw the emergence of AI, just like other industries. In a dental clinic, it can carry out simple and difficult tasks with higher precision, accuracy, sensitivity, and—most importantly—in less time [10]. In recent years, Machine Learning algorithms have the potential to be used to develop automated caries detection systems that are more accurate and efficient than traditional methods. Adaptive neural network architecture [11], deep learning [12], an artificial multilayer perceptron neural network [13], convolutional neural network [14], backpropagation neural network [15], and

\*Corresponding Author.

k-means clustering [16] are some of the different methods used in dentistry, specifically for the detection of caries. A large, difficult assignment has been seen to disappear utilizing these strategies. Therefore, this review aims to provide an overview of the diverse artificial techniques used to identify dental cavities in this systematic review, as follows:

**B. Artificial Neural Networks**

Artificial Neural Networks (ANNs) are computational systems profoundly influenced by the functioning of biological nervous systems, such as the human brain. ANNs consist predominantly of numerous interconnected computational units, commonly known as neurons. These neurons collaboratively operate in a distributed manner to assimilate knowledge from input data, with the objective of refining their ultimate output.

O’Shea [17] represented the fundamental architecture of an ANN as depicted in Fig. 1.

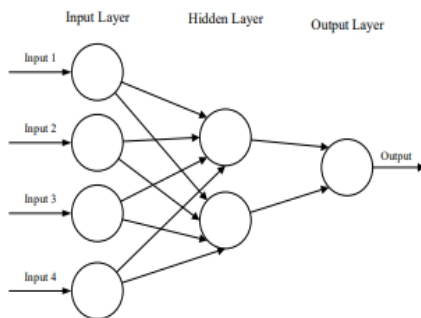


Fig. 1. The basis of a number of common ANN architectures.

The input, often organized as a multidimensional vector, is introduced to the initial layer, known as the input layer. Subsequently, this input is propagated through intermediary layers called hidden layers. In these hidden layers, decisions are made based on the preceding layer’s information. The hidden layers then assess how alterations within themselves positively or negatively impact the final output. This iterative process of evaluating and adjusting is termed learning. The stacking of multiple hidden layers, creating a tiered arrangement, is commonly referred to as deep learning.

**C. Adaptive Neural Network Architecture**

An adaptive neural network architecture refers to a type of artificial neural network that can dynamically adjust its structure and parameters based on the characteristics of the input data. In the context of images, an adaptive neural network architecture is designed to intelligently adapt its layers, nodes, or connections to better capture the features present in the input image. This concept aims to enhance the network’s performance by tailoring its architecture to the specific complexities and patterns within the image data.

The adaptive nature of such architectures allows the neural network to optimize its internal representation as it learns from the data. Traditional neural network architectures have fixed structures, making them less flexible in handling variations in data characteristics. In contrast, an adaptive neural network architecture has the ability to modify itself during training, potentially leading to improved accuracy, efficiency, and

generalization. Haykin [18] presents a conceptual framework outlining a singular stage of neural processing intended for adaptable behavior in Fig. 2.

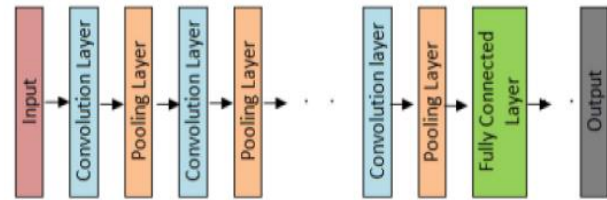


Fig. 2. Schematic diagram of an adaptive system.

The model centers on the retention of past experiences to predict likely future occurrences. Specifically, for a given input vector  $x(n-1)$  at time  $n-1$ , the model estimates the expected value  $x^*(n)$  at time  $n$ . By comparing this prediction to the actual value  $x(n)$ , the difference, termed the correction or innovation signal, is computed. A non-zero correction signifies an unfamiliar condition, necessitating model updates to better anticipate similar situations. This dynamic adjustment enables the model to learn and adapt to its environment, as it continually operates in real-world scenarios.

**D. Convolutional Neural Network**

A convolutional neural network (CNN) represents an evolved variant of artificial neural networks, meticulously designed to handle the intricate analysis of visual data like images and videos. Fig. 3 shows the concept model of convolutional neural network. Drawing inspiration from the intricate visual processing mechanism observed in the human brain, CNNs demonstrate exceptional proficiency in tasks that encompass image recognition, classification, and the domain of computer vision. The quintessential prowess of CNNs originates from their innate ability to independently glean intricate features from visual content, thereby accentuating their utility in intricate data interpretation. This is achieved through several distinctive architectural components:

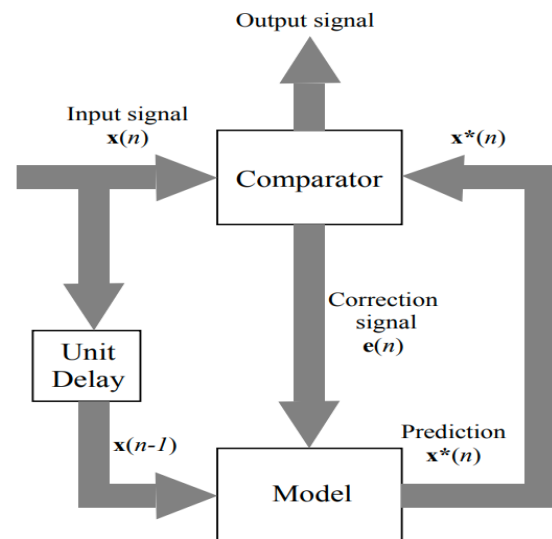


Fig. 3. The concept model of convolutional neural network [19].



- **Convolutional Layers:** These layers employ adaptable filters for performing convolution operations on input images. This hierarchical feature extraction encompasses both rudimentary features, such as edges, and more intricate features, like object contours.
- **Pooling Layers:** Subsequent to convolution, pooling layers downsize the spatial dimensions of features while retaining vital information. Employing methods like max-pooling, these layers preserve the maximum value within localized regions, distilling key features. This pooling process bolsters resilience against minor input variations and concurrently reduces computational requirements.
- **Fully Connected Layers:** Features extracted from prior layers are flattened and channeled through fully connected layers. These layers mirror conventional neural network structures and undertake roles as classifiers or regressors, relying on the features they have learned.

**E. Backpropagation Neural Network**

The backpropagation algorithm, integral to neural network training [20], operates as a pivotal mechanism in optimizing model parameters for improved performance in tasks such as classification, regression, and pattern recognition. This process empowers neural networks to glean insights from labeled training data, effectuating adjustments in weights and biases to minimize the disparity between projected outputs and actual target values. Backpropagation fosters neural networks to refine parameters, facilitating an enhanced fit to training data. By iteratively adjusting weights and biases using gradients, networks learn to identify pertinent features and relationships within data. This systematic learning process enables neural networks, encompassing deep learning architectures like CNNs and Recurrent Neural Networks (RNNs), to achieve elevated levels of performance across various tasks.

**F. K-Means Clustering**

K-means clustering is a widely utilized unsupervised machine learning technique employed to partition a dataset into distinct groups, or clusters, based on inherent patterns in the data. This technique is particularly effective in uncovering underlying structures and relationships within unlabeled datasets. The K in K-means represents the user-defined number of clusters. The choice of K significantly impacts the quality of the clustering results.

**G. Studies of Predicting Depression based on Self-Supervised Learning Method**

Self-Supervised Learning is an emerging machine learning paradigm that leverages unlabeled data to train models in a semi-supervised manner [21]. A two-phase learning scheme in self-supervised learning is illustrated by Taleb et al. [22]. Fig. 4 shows the flowchart of self-supervised learning stages.

Unlike traditional supervised learning, where labeled data is used to directly predict specific targets, self-supervised learning formulates tasks that allow the model to learn meaningful representations from the data itself. In self-supervised learning, the learning signal comes from the data

itself, generating surrogate tasks that help the model capture underlying patterns and structures. For a more comprehensive perspective on self-supervised learning, we shall conduct a comparative examination juxtaposed against alternative machine learning techniques. Several studies are illustrated in Table I.

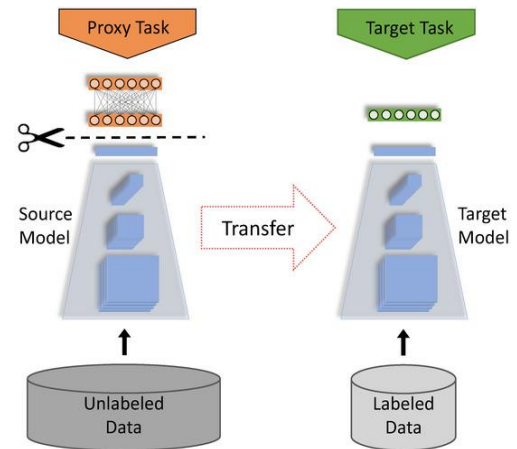


Fig. 4. Flowchart of self-supervised learning stages.

TABLE I. SUMMARY OF DENTAL CARIES DETECTING STUDIES

Article	Data	Models /Algorithms	Results
Patil et al. (2019)[11]	Small size of dataset (45 dental images)	K-nearest neighbor, ANN	Accuracy=95%, Precision=90%
Casalegno et al. (2019)[12]	217 X-dental images	Convolutional Neural Network (CNN) algorithm	Accuracy of 85.6% and 83.6%
Devito et al. (2008)[14]	160 radiographic images	Backpropagation algorithm	Accuracy= 88.4%
Zanella-Calzada et al. (2018)[23]	9812 subjects were in an age range of 0 to 80 years old; 4830 belonged to the masculine gender and 4982 to the feminine gender	ANN by classifying subjects	Accuracy= 88%
Lee et al. (2018)[24]	2417 images (853 healthy tooth surfaces/1,086 non-cavitated carious lesions/431 cavitations/47 automatically excluded images during preprocessing).	Convolutional Neural Networks	Accuracy= 93%
Taleb et al. (2022)[22]	38,094 bitewing radiographs	Self-Supervised Learning Algorithms	ROC-AUC= 71.50 Sensitivity=51.80 Specificity of 91.30

### III. RESULTS AND DISCUSSION

Patil et al. (2019) [11] embarked on a comprehensive exploration to discern the optimal algorithm for diagnosing tooth caries. This endeavor entailed a thorough assessment of multiple algorithms, including support vector machines (SVM), k-nearest neighbors (KNN), Naïve Bayes (NB), and the adaptive dragonfly algorithm (ADA-NN). Their dataset encompassed 120 dental images, forming the foundation for a series of three distinct tests, each centered around 40 dental images. Notably, the ADA-neural network emerged as a consistent frontrunner across these evaluations, showcasing superior performance vis-à-vis the aforementioned algorithms. Impressively, the ADA-neural network surpassed its counterparts by margins of 5.5%, 11.76%, and 6.5%, respectively. However, it's crucial to acknowledge that Patil and collaborators operated within a notably constrained dataset, comprising a mere 45 images, thereby raising legitimate queries regarding the reliability of their findings. The utilization of such a limited dataset, consisting of merely 45 images, inevitably instills doubts concerning the robustness of their outcomes. Within this subset, 30 images were earmarked for training, while the remaining 16 were allocated for testing.

In parallel, the studies orchestrated by Devito et al. [14] harnessed the potent backpropagation algorithm within deep learning for dental caries prognosis. In the realm of Devito et al.'s investigation, the training dataset encompassed 80 dental images, with the remaining 80 images partitioned into two distinct subsets: 40 images for validation purposes and an additional 40 images designated for comprehensive testing [14]. Expanding the scope of inquiry, Devito and associates directed their efforts toward forecasting the proximal category of dental caries through the prism of X-ray dental radiographs, culminating in a commendable accuracy level of 88.4%.

Likewise, Casalegno et al. (2019) [12] pursued a separate avenue of exploration, deploying a dataset brimming with 217 X-ray dental images. Embracing the CNN algorithm, their study undertook the ambitious task of caries prediction, encompassing the nuanced realms of proximal and occlusal caries. Impressively, the outcomes bore witness to a level of accuracy amounting to 85.6% for proximal caries and 83.6% for occlusal caries.

The expedition orchestrated by Lee et al. (2018) [24] unfolded within the realm of CNNs, their dataset comprising 2,417 images. This comprehensive assemblage featured 853 images depicting healthy tooth surfaces, 1,086 images of non-cavitated carious lesions, 431 images capturing cavitations, and a subset of 47 images that underwent automatic exclusion during the preprocessing phase. It's noteworthy; however, Lee and co-authors deviated from the conventional dataset division percentages of 25%, 50%, 75%, and 100% for training. Rather, their dataset was bifurcated into a training set (comprising 1,891/673/870/348 images for each respective category) and a test set (consisting of 479/180/216/83 images for the corresponding categories). This unconventional distribution, though divergent, did not deter them from achieving an impressive diagnostic precision through CNNs, boasting an accuracy rate approximating 93.3%.

Taleb et al. (2022) [22] utilized a Self-Supervised Learning Algorithms on dental caries detection. The dataset was obtained by three specialized dental clinics in Brazil, focusing on radiographic and tomographic examinations [22] and consisted of 38,094 BWRs taken between 2018 and 2021. The study's strengths included its pioneering demonstration of self-supervised techniques in dentistry, with the potential to address the immense volume of X-ray images generated globally. Additionally, it employed a dataset of over 30,000 Bitewing Radiographs (BWRs) with EHR-based labels, overcoming the challenge of diagnostic inconsistency by incorporating a refined ground truth. Nonetheless, limitations included the use of EHR-based labels, which may be biased and incomplete, and the focus on tooth-level classification rather than finer assessments. This discrepancy might account for the relatively minor decrement in study results compared to those reported by Lee et al. [24].

In summary, the amalgamation of Self-Supervised Learning Algorithms has shown effectiveness in enhance performance and optimize label utilization in scenarios involving dental caries classification. Nevertheless, the predictive efficacy of machine learning approaches differs between studies due to disparities in data distribution, the characteristics of features integrated into the model, and the manner in which the outcome variable is gauged. Consequently, while certain investigations have indicated proficient performance of ML algorithms, a persistent requirement remains for further research to authenticate the predictive capabilities of each algorithm. This necessity arises from the inability to universally extend the results to encompass all forms of data.

### IV. CONCLUSION

In conclusion, the potential of self-supervised learning for advancing dental caries detection is substantial. By pretraining models on large datasets derived from routine care, self-supervised learning provides a practical solution for scenarios where labeled data is limited. The presented studies underscore the positive impact of self-supervised learning algorithms on the predictive performance of dental caries classification models. However, it's important to acknowledge that the success of self-supervised learning is not uniform across all scenarios. Variability in data distribution, feature characteristics, and outcome measurement can lead to differing predictive performances. While some studies have demonstrated impressive results, the applicability of these findings to all types of data requires careful consideration.

Despite these challenges, the prospect of self-supervised learning remains promising. It offers a pathway to leverage the vast amounts of unannotated data generated in routine clinical practice, potentially revolutionizing dental diagnostics. As this field continues to evolve, further research is imperative to refine methodologies, validate findings, and establish the generalizability of self-supervised learning techniques in dental caries detection. The integration of self-supervised learning into clinical practice could mark a significant advancement in early caries diagnosis, ultimately leading to improved patient care and oral health outcomes.

#### 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] Isabella Castiglioni, Leonardo Rundo, Marina Codari, Giovanni Di Leo, Christian Salvatore, Matteo Interlenghi, Francesca Gallivanone, Andrea Cozzi, Natascha Claudia D'Amico, Francesco Sardanelli, AI applications to medical images: From machine learning to deep learning. *Physica Medica*, vol. 83, pp. 9-24, 2021
- [2] Alexandra T. Greenhill, Bethany R. Edmunds, A primer of artificial intelligence in medicine. *Techniques and Innovations in Gastrointestinal Endoscopy*, vol 22, issue 2, pp. 85-89, 2020.
- [3] A. Rajkomar, J. Dean, I. Kohane, Machine Learning in Medicine. *Adv Exp Med Biol*, N Engl J Med, pp. 1347-1358, 2019.
- [4] Stefano A. Bini, Artificial Intelligence, Machine Learning, Deep Learning, and Cognitive Computing: What Do These Terms Mean and How Will They Impact Health Care? *The Journal of Arthroplasty*, vol. 33, issue 8, pp. 2358-2361, 2018.
- [5] Grünberg, K., Jimenez-del-Toro, O., Jakab, A., Langs, G., Salas Fernandez, T., Winterstein, M., ... & Krenn, M. (2017). Annotating medical image data. *Cloud-Based Benchmarking of Medical Image Analysis*, pp45-67.
- [6] Ericsson, L., Gouk, H., Loy, C. C., & Hospedales, T. M. Self-supervised representation learning: Introduction, advances, and challenges. *IEEE Signal Processing Magazine*, vol 39(3), pp. 42-62, 2022.
- [7] Nima Tajbakhsh, Laura Jeyaseelan, Qian Li, Jeffrey N. Chiang, Zhihao Wu, Xiaowei Ding. Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. *Medical Image Analysis*, vol 63, pp. 101693, 2023
- [8] Xu, J. A Review of Self-supervised Learning Methods in the Field of Medical Image Analysis. *Int. J. Image Graph. Signal Process.* vol 13, pp. 33-46, 2021 .
- [9] Selwitz, R. H., Ismail, A. I., & Pitts, N. B. Dental caries. *The Lancet*, issue 9555, vol 369(9555), pp. 51-59, 2007.
- [10] N. Ahmed, M. S. Abbasi, F. Zuberi et al., Artificial intelligence techniques: analysis, application, and outcome in dentistry-A systematic review. *BioMed Research International*, vol. 2021, pp. 1-15, 2021.
- [11] S. Patil, V. Kulkarni, and A. Bhise. Algorithmic analysis for dental caries detection using an adaptive neural network architecture. *Heliyon*, vol. 5, no. 5, 2019.
- [12] F. Casalegno et al. Caries detection with near-infrared transillumination using deep learning, *Journal of Dental Research*, vol. 98, no. 11, pp. 1227-1233, 2019.
- [13] C. S. Mackenzie, W. L. Gekoski, and V. J. Knox. Age, gender, and the underutilization of mental health services: the influence of help-seeking attitudes. *Aging Ment Health*, vol. 10, no. 6, pp. 574-582, 2006.
- [14] K. L. Devito, F. de Souza Barbosa, and W. N. F. Filho. An artificial multilayer perceptron neural network for diagnosis of proximal dental caries, *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology & Endodontics*, vol. 106, no. 6, pp. 879-884, 2008.
- [15] F. Schwendicke, T. Golla, M. Dreher, and J. Krois. Convolutional neural networks for dental image diagnostics: a scoping review, *Journal of Dentistry*, vol. 91, p. 103226, 2019.
- [16] V. Geetha, K. S. Aprameya, and D. M. Hinduja. Dental caries diagnosis in digital radiographs using back-propagation neural network, *Health Information Science and Systems*, vol. 8, no. 1, p. 8, 2020.
- [17] O'Shea, K., & Nash, R. An introduction to convolutional neural networks. arXiv preprint arXiv:1511.08458, 2015..
- [18] Haykin, Simon. *Neural Networks A Comprehensive Foundation*. New York: Macmillan College Publishing Company, 1994.
- [19] F. Sultana, A. Sufian, and P. Dutta. Advancements in image classification using convolutional neural network. In 2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), pp. 122-129, Nov 2018.
- [20] Zajmi, L., Ahmed, F. Y., & Jaharadak, A. A. Concepts, methods, and performances of particle swarm optimization, backpropagation, and neural networks. *Applied Computational Intelligence and Soft Computing*, 2018
- [21] L. Jing and Y. Tian. Self-supervised visual feature learning with deep neural networks: A survey. *IEEE TPAMI*, 2020.
- [22] Taleb A, Rohrer C, Bergner B, De Leon G, Rodrigues JA, Schwendicke F, Lippert C, Krois J. Self-Supervised Learning Methods for Label-Efficient Dental Caries Classification. *Diagnostics*. 2022;
- [23] L. Zanella-Calzada, C. Galván-Tejada, N. Chávez-Lamas et al., "Deep artificial neural networks for the diagnostic of caries using socioeconomic and nutritional features as determinants: data from NHANES 2013-2014," *Bioengineering*, vol. 5, no. 2, pp. 47, 2018.
- [24] J.-H. Lee, D.-H. Kim, S.-N. Jeong, and S.-H. Choi, "Detection and diagnosis of dental caries using a deep learning-based convolutional neural network algorithm," *Journal of Dentistry*, vol. 77, pp. 106-111, 2018.

# Optimization Method for Trajectory Data Based on Satellite Doppler Velocimetry

Junzhuo Li<sup>1\*</sup>, Wenyong Li<sup>2</sup>, Guan Lian<sup>3</sup>

School of Information and Communication, Guilin University of Electronic Technology, Guilin 541004, China<sup>1,2</sup>  
Guilin University of Electronic Technology Guangxi Key Laboratory of Intelligent Transportation, Guilin 541004, China<sup>3</sup>

**Abstract**—Due to cost and energy consumption limitations, there are significant differences in the positioning capabilities of mobile terminals, resulting in unsatisfactory quality of trajectory data. In this paper, satellite Doppler data is used to optimize trajectory data. First, the system state equation is established by the kinematic relationship between the measured velocity and position, and the static linear Kalman filter estimates the optimal system state. Then a dynamic Kalman filter system is established by correlating the measurement error matrix parameters of the Kalman filter with the vertical dilution of precision of satellite positioning. Finally, the whole-day trajectory of a taxi in Shenzhen was visualized, and the deviation between the trajectory points and the urban road was calculated to compare the optimized and non-optimized taxi trajectories. The results show that the proposed optimization method can effectively reduce the deviation between trajectory points and urban roads, and this method can be used to process vehicle trajectory data in urban traffic research.

**Keywords**—Urban transportation; Kalman Filter; information fusion; trajectory data

## I. INTRODUCTION

The rise of mobile Internet and location-based services (LBS) has generated a large amount of trajectory data. These data provide abundant materials for analyzing and researching social behavior and economic activities and have significant scientific research value. In the study of social phenomena, trajectory data have many advantages [1, 2], such as: 1) Trajectory data is an objective record and description of the position changes of mobile terminals and users, which is not affected by subjective feelings and has objectivity. 2) Mobile Internet has the characteristics of timeliness, which makes trajectory data update very fast and can provide timely feedback on social phenomenon and behavior changes. 3) With the popularity of mobile terminals such as mobile phones, location-based services have been widely applied, generating a massive amount of trajectory data. 4) Mobile terminals with positioning functions, such as mobile phones, tablets, taxis, and shared bicycles, can generate various trajectory data through various methods such as GPS, cellular network positioning, WI-FI positioning, etc.

During the collection process of trajectory data, there are inevitably systematic errors due to positioning technology. For example, the positioning accuracy of GSP is related to factors such as the signal reception ability of the receiver, calculation methods, local shading conditions, atmospheric conditions, and the number of observable satellites, and the positioning accuracy of cellular networks is closely related to network

format, solution methods, and the density of signal base stations. Due to cost, energy consumption, safety, and other reasons, mobile terminals such as mobile phones can usually only receive L1 carrier signals and C/A codes and cannot use differential positioning based on carrier phase. Therefore, optimizing their trajectories is very crucial [3].

Improving positioning and information communication technology is a direct way to improve positioning accuracy. For example, the third-generation GPS satellite launched in 2018 can launch a new civilian signal L1C in the 1575.42MHz band, making GPS more compatible with other GNSS systems and enabling receivers to receive more satellite signals [4]. In addition, the third-generation GPS satellite is equipped with a better accurate rubidium atomic clock and has higher signal transmission power [5,6]. With the upgrading of communication technology, the accuracy of positioning technology based on cellular networks is also unceasingly improving [7]. In 3GPP Release 16, a new Down Link-Position Reference (DL-PRS) is brought in, and various positioning technologies with better performance, such as DL-TDOA, UL-TDOA, DL-AOD, UL-AOA, E-CID, could be used in the fifth-generation communication network [8]. In the fifth-generation communication network, the dense network makes TDOA and DOA positioning more reliable, Massive Multiple Input Multiple Output (Massive MIMO) technology provides an Infrastructure for AOA positioning, and lower network latency improves the accuracy of time-based positioning methods [9-11].

In addition to improvements in positioning and communication technology, data processing and fusion can also improve the quality of trajectory data. Using spatial clustering and filtering algorithms can make trajectory data smoother, avoiding deviations or anomalies in trajectory data [12-15]. By integrating information beyond positioning data, such as electronic maps, Doppler velocimetry, and vehicle inertial measurement units, the quality of trajectory data can also be improved [16-18].

In this study, the trajectory data is optimized using the velocity measured by GPS satellites based on the Doppler effect. The structure of this article is as follows: In Section II, the system state equation is established according to the kinematics principle, and then the acquisition of the measurement value of the system state equation is introduced, that is, the pseudo-range positioning method and the Doppler velocity measurement method of the satellite. In Section III, the static Kalman filter is used to estimate the optimal system state. Then the dynamic Kalman filter is used to estimate the

optimal system state by correlating the system measurement matrix with the satellite positioning Dilution of Precision (DOP). In Section IV, the daily trajectory of a taxi in Shenzhen was visualized, and the deviation between the trajectory and the urban road network was calculated to compare the optimized and non-optimized taxi trajectories.

## II. ESTABLISH SYSTEM STATE EQUATION

The state equation of the Kalman filter system mainly includes two prediction and measurement processes. According to the laws of kinematics, the system state equation can be obtained as Formula (1). In this formula,  $o_k$  is the system state at time  $k$ , which includes  $(x^k, y^k)$  meaning the position of the object and  $(v_x^k, v_y^k)$  meaning the speed of the object.  $A_k$  is the system state transition matrix,  $w_k$  is the system noise, which conforms to the Gaussian distribution with the mean value of 0, and the variance of  $Q$ ,  $v_k$  is the measured noise, which conforms to the normal distribution with the mean value of 0 and the variance of  $R$ .

$$\begin{cases} o_k = A_k x_{k-1} + w_k \\ z_k = o_k + v_k \end{cases} = \begin{bmatrix} x^k \\ y^k \\ v_x^k \\ v_y^k \\ z_k = o_k + N \sim (0, R) \end{bmatrix} = \begin{bmatrix} 1 & \Delta t & & & \\ & 1 & \Delta t & & \\ & & 1 & & \\ & & & 1 & \\ & & & & 1 \end{bmatrix} \begin{bmatrix} x^{k-1} \\ y^{k-1} \\ v_x^{k-1} \\ v_y^{k-1} \end{bmatrix} + N \sim (0, Q) \quad (1)$$

### A. System Measurement: GPS Pseudo-Range Positioning

GPS has the advantages of fast positioning speed, accurate positioning, low cost, and wide coverage. The positioning results can be obtained by solving Formula (2) with four unknown variables  $x$ ,  $y$ ,  $z$ ,  $\delta$ . When the receiver can receive four satellite signals, the equation has a unique solution; when the received signal is less than 4, the GPS satellite cannot independently calculate the receiver position; When the received signals are more than 4, the least square method is usually used to solve the overdetermined equations. When many satellite signals can be received, the more the number of equations, the more stable the solution result, and the more minor effect of one satellite measurement deviation on the positioning result, making the positioning result more accurate and stable [19, 20]. The positioning accuracy of GPS is closely related to the number of satellite signals received. Due to cost and energy consumption limitations, mobile terminals' satellite signal reception capacity is commonly insufficient, and high-rise buildings could block GPS signal transmission in urban areas. Many reasons will affect the reception of satellite signals, resulting in inaccurate positioning, so improving the accuracy of trajectory data is crucial [21].

$$(x^s - x)^2 + (y^s - y)^2 + (z^s - z)^2 = (\rho^s - \delta c)^2 \quad (2)$$

In Formula (2),  $[x^s, y^s, z^s]$  is the coordinates of the received satellite;  $[x, y, z]$  is the coordinates of the receiver;  $\delta$  is the receiver's clock error;  $c$  is the speed of light;  $\rho^s$  is the pseudo-range of the received satellite, which can be calculated by Formula (3).

$$\rho^s = (r^s + \delta - \delta^s + I + T)c + \varepsilon^s \quad (3)$$

In Formula (3),  $r^s$  is the time required for the signal to be transmission between the receiver and the satellite under vacuum conditions,  $\delta^s$  is the clock error of the satellite,  $I$  is the ionospheric delay,  $T$  is the tropospheric delay, and  $\varepsilon^s$  is the error of the pseudo-range.

### B. System Measurement: Satellite Doppler Velocimetry

Formula (3) calculates the derivative of time to obtain Formula (4).

$$\dot{\rho}^s = (\dot{r}^s + \dot{\delta} - \dot{\delta}^s + \dot{I} + \dot{T})c + \dot{\varepsilon}^s \quad (4)$$

In Formula (4),  $\dot{\rho}^s$  is the rate of change of pseudo-range, which can be calculated by Doppler frequency shift [22], as in Formula (5);  $\dot{r}^s c$  is the change rate of the geometric distance between the receiver and satellite, as shown in Formula (6);  $\dot{\delta}$  and  $\dot{\delta}^s$  are frequency drift of receiver and satellite;  $\dot{I}$  and  $\dot{T}$  are the change rate of ionospheric delay and tropospheric delay, which are small enough to be ignored [23].

$$\dot{\rho}^s = -\lambda(f - f^s) \quad (5)$$

In Formula (5),  $\lambda$  is the wavelength of the transmitted signal,  $f$  and  $f^s$  are the frequency of the signal received by the receiver and the frequency of the signal transmitted by satellite, respectively.

$$\dot{r}^s c = (v^s - v)I^s \quad (6)$$

In Formula (6),  $v^s = [v_x^s, v_y^s, v_z^s]^T$  is the travel speed of the satellite;  $v = [v_x, v_y, v_z]^T$  is the travel speed of the terminal device or user, which is an unknown variable to be solved;  $I^s$  is the directional unit vector of the receiver, which can be calculated according to Formula (7).

$$I^s = \frac{1}{\sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} \quad (7)$$

Substitute Formula (5-6) into Formula (4) to get Formula (8), and then sort out all unknown variables into the left side to get Formula (9) with four unknown variables  $v_x$ ,  $v_y$ ,  $v_z$ ,  $\dot{\delta}$ . When the number of satellite signals that can be received is greater than or equal to four, the travel speed of the mobile terminal or user can be obtained by solving Formula (9).

$$-\lambda(f - f^s) = (v^s - v)I^s + \dot{\delta} - \dot{\delta}^s + \dot{\varepsilon}^s \quad (8)$$

$$vI^s - \dot{\delta} = \lambda f - \lambda f^s + v^s I - \delta^s + \varepsilon^s \quad (9)$$

### III. KALMAN FILTERING SYSTEM

Kalman filter is an optimal estimation method using the system state equation. When the measurement variance is known, the Kalman filter can estimate the optimal state of the system from a series of measurements containing noise [24, 25]. This Section introduces the basic calculation process of the Kalman filter, and the static Kalman filter method is used to calculate the changing trend of system Kalman gain and system state uncertainty. Then, the random error of the measured value of the system is correlated with the DOP of satellite positioning to establish a dynamic Kalman filter system.

#### C. Static Kalman Filter System

The linear Kalman filter method calculation includes two stages of prediction and update involving five formulas.

In the prediction stage, the calculation of the linear Kalman filter includes:

- System state prediction equation. It uses the previous state  $\hat{\sigma}_{k-1}$  to infer the current state  $\hat{\sigma}_k$ , such as Formula (10). In Formula (10),  $A$  is the state transition matrix, which establishes the relationship between the current system state and the system state at the previous moment through kinematics laws.

$$-\lambda(f - f^s) = (v^s - v)I^s + \dot{\delta} - \delta^s + \varepsilon^s \quad (10)$$

- Prediction equation of system uncertainty. It calculates the prior estimate  $P_k$  of the current system state uncertainty according to the uncertainty  $P_{k-1}$  of the previous moment and the system's random error  $Q$  of the observation value, as shown in Formula (11).

$$-\lambda(f - f^s) = (v^s - v)I^s + \dot{\delta} - \delta^s + \varepsilon^s \quad (11)$$

- In the update stage, the calculation of the linear Kalman filter includes:

$$K_k = \frac{P_k}{P_k + R} \quad (12)$$

- System state update equation. It calculates the current system optimal state  $\hat{\sigma}_k$  according to the prior estimate  $\hat{\sigma}_k$  of the system state, the current measured value  $z_k$ , and the Kalman gain  $K_k$ , such as Formula (13).

$$\hat{\sigma}_k = \hat{\sigma}_k + K_k(z_k - \hat{\sigma}_k) \quad (13)$$

- Update equation of system uncertainty. It calculates the optimal uncertainty  $P_k$  of the current system state according to the prior estimated uncertainty  $P_k$  and Kalman gain  $K_k$ , such as Formula (14). In Formula (14),  $E$  is an identity matrix.

$$P_k = (E - K_k)P_k \quad (14)$$

The calculation flow of the whole Kalman filter is shown in Fig. 1. The upper level of Fig. 1 includes two initial and measurement modules responsible for data input. The initialization module inputs the iterative initial values of  $\hat{\sigma}_0$  and  $P_0$ , and the parameters needed for random error  $R$ , systematic error  $Q$ , and transition matrix  $A_k$  to establish the system equation. The measurement module continuously inputs the measured values  $z_k$  in the iterative calculation process of the system. The middle level of Fig. 1 is the core of the calculation in the Kallman filter, which uses Formula (10-14) to iterate the system's state, including the prediction and update modules. The lower level of Fig. 1 is responsible for the output of the system's optimal state and uncertainty.

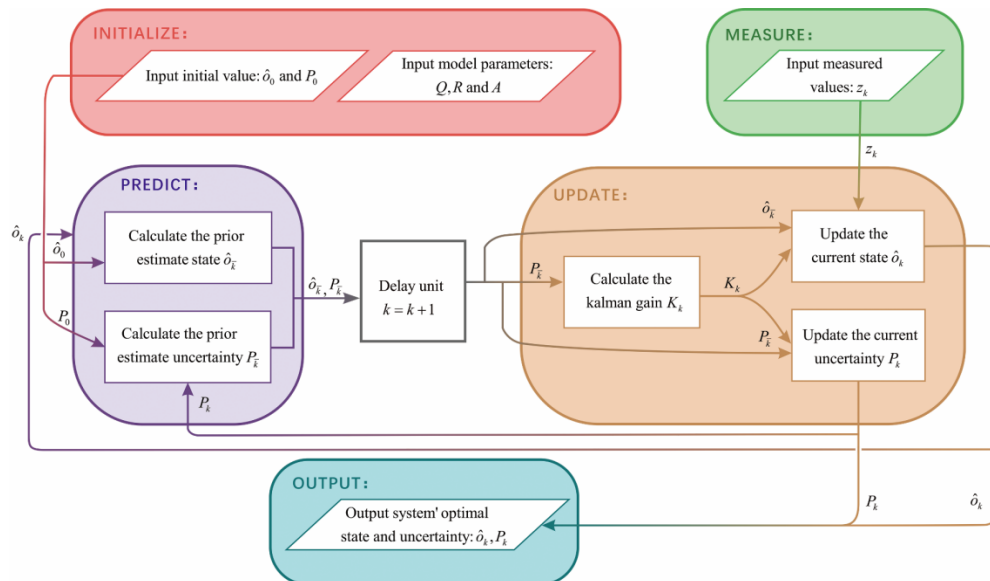


Fig. 1. The calculation flow of the linear kallman filter.

This study uses the Doppler velocity measurement and pseudo-range positioning results of GPS to establish the system state, with a calculation step of  $\Delta t=10$ , and the settings of the system error and random error of measurement are shown in Formula (15-16).

$$Q = \begin{bmatrix} \varepsilon_x & \varepsilon_y & \varepsilon_{vx} & \varepsilon_{vy} \\ 0.5 & & & \\ & 0.5 & & \\ & & 0.1 & 0.1 \\ & & 0.1 & 0.1 \end{bmatrix} \begin{matrix} \varepsilon_x \\ \varepsilon_y \\ \varepsilon_{vx} \\ \varepsilon_{vy} \end{matrix} \quad (15)$$

$$R = \begin{bmatrix} \partial_x & \partial_y & \partial_{vx} & \partial_{vy} \\ 2.5 & & & \\ & 2.5 & & \\ & & 0.5 & 0.5 \\ & & 0.5 & 0.5 \end{bmatrix} \begin{matrix} \partial_x \\ \partial_y \\ \partial_{vx} \\ \partial_{vy} \end{matrix} \quad (16)$$

According to the established system state equation, the variation trend of the Kalman gain and uncertainty with iteration is calculated and plotting Fig. 2. This Kalman filter system is static because the parameters  $A$ ,  $Q$ , and  $R$  of the system are constant. According to Fig. 2, the Kalman gain and uncertainty gradually decrease and converge with the calculation iteration in the static system.

#### D. Dynamic Kalman Filtering System

The transition matrix  $A$  is invariant, determined by the kinematic relationship between position and speed in the system state equation, as well as the system error matrix  $Q$  will not be easily changed in specific equipment. However, the number and geometric distribution of received satellites in space change over time according to the ephemeris and can affect the measurement random error  $R$ . When  $R(t)$  is added to the system instead of  $R$ , a dynamic Kalman filter system can be obtained [26]. Generally, GPS measurement random error can be indicated by the DOP. Fig. 3

shows the DOP and the number of received satellites in the study area on the day of the experiment.

Fig. 3 shows five different DOPs, geometric DOP, time DOP, position DOP, horizontal DOP, and vertical DOP, among which the horizontal DOP is the most suitable for establishing a functional relationship with  $R(t)$ . At 9:50, the maximum value of horizontal DOP was 1.66, and at 00:00, the minimum value of horizontal DOP was 0.73. Therefore, the variation trend of the Kalman gain and system state uncertainty with iteration was calculated when  $R(t)=1.66$  and  $R(t)=0.73$ , as shown in Fig. 4. In the dynamic Kalman filter system, the Kalman gain and the system state uncertainty will vary within the values of the filling area in Fig. 4. With the iterative calculation of the system, the Kalman gain and the system state uncertainty will gradually stabilize and fluctuate in a small range.

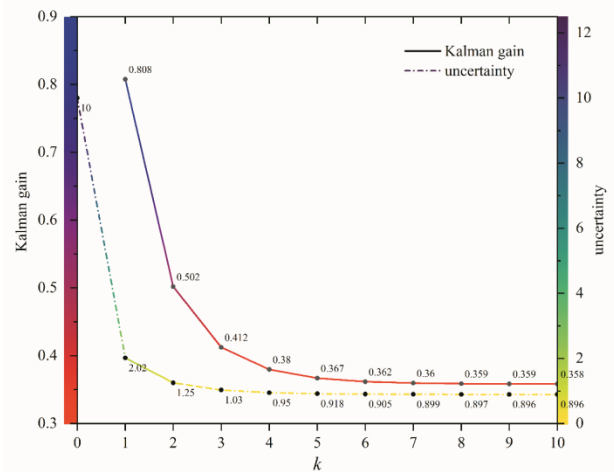


Fig. 2. The variation trend of the kalman gain and uncertainty in the iterative calculation.

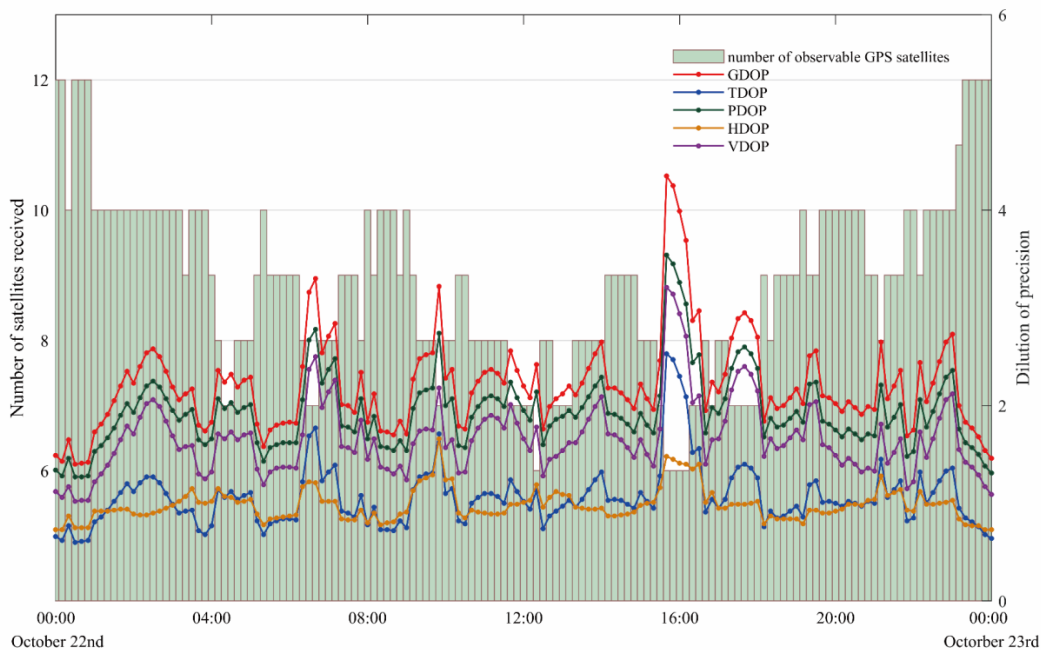


Fig. 3. The DOP and the number of received satellites in the study area on the day of the experiment.

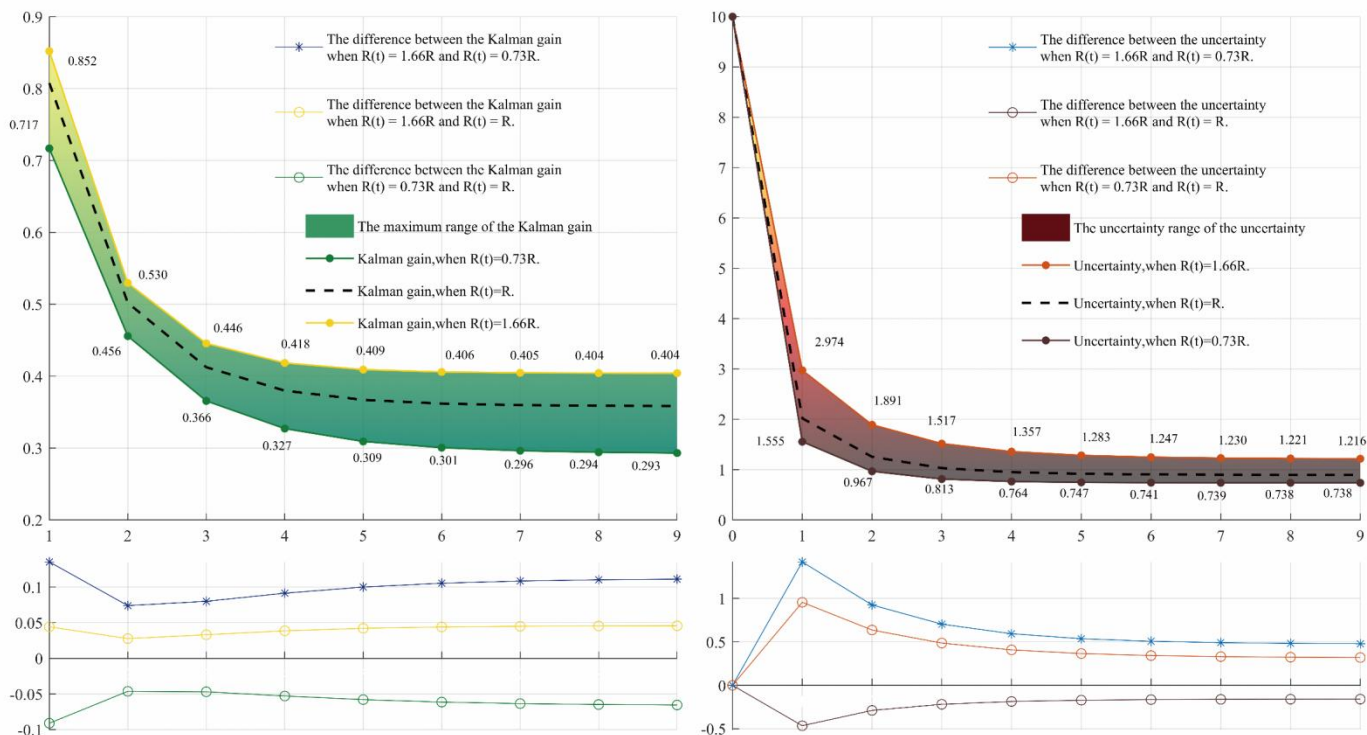


Fig. 4. The variation range of the kalman gain and the uncertainty in the dynamic kalman filtering system.

#### IV. EXPERIMENT

##### A. Study Region

In this study, the performance of the proposed method is verified by using the taxi travel trajectory data in Shenzhen.

The longitude range of the study area is 113.8 to 114.25, and the latitude range is 22.5 to 22.75. The study date is October 22, 2013. Fig. 5 shows the study area's administrative divisions, natural resources, land use, and urban road conditions.

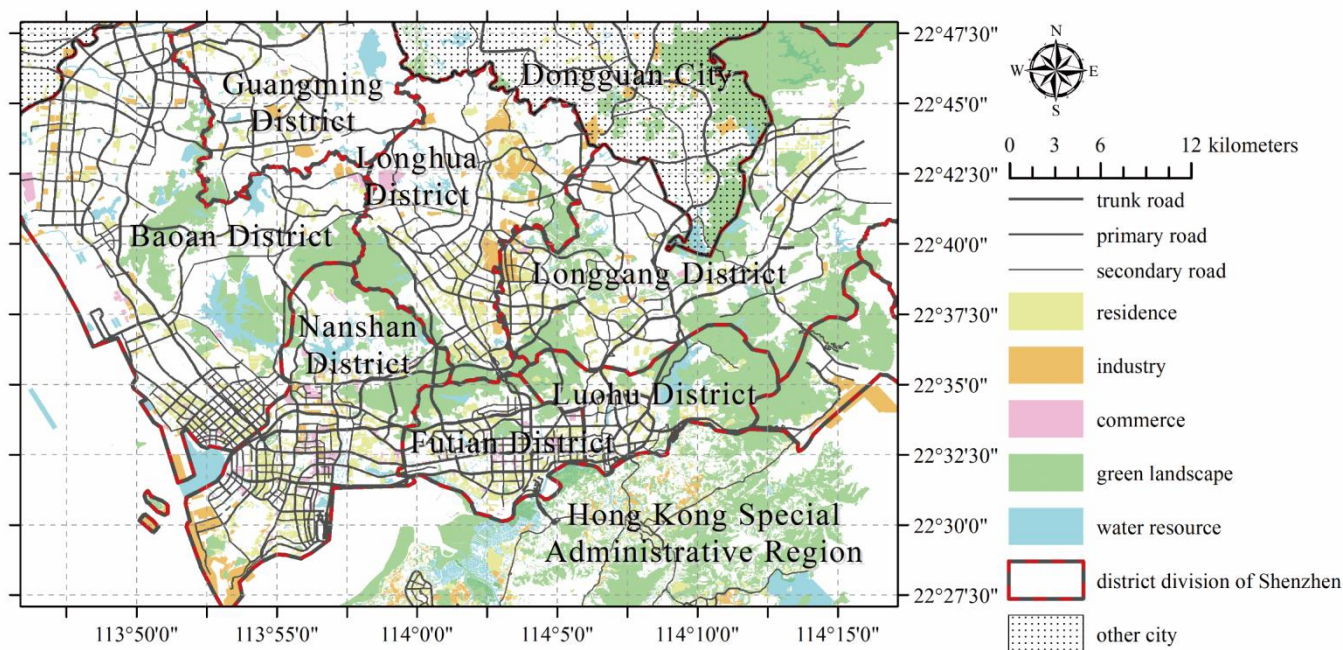


Fig. 5. The geographic information of the study region.



### B. Data Description

The data set records the taxi ID, current time, longitude, latitude, speed, and occupancy status. Table I shows some data.

TABLE I. EXAMPLES OF SOME RECORDS IN THE DATA SET

taxi ID	time	longitude	latitude	speed (m/s)	occupancy status
22223	00:03:39	114.167732	22.562550	2.56	1
22223	00:03:54	114.168999	22.562550	11.94	1
22223	00:04:09	114.170998	22.562550	13.06	1
22223	00:04:23	114.172897	22.562599	15.11	1
22223	00:04:39	114.175232	22.561701	18.33	1

The field taxi ID is used to distinguish between different taxis and has no practical meaning. The time, longitude, latitude, and speed fields record an equipment's current time, position, and speed. The field occupancy status records whether the taxi has an ongoing order, with 1 meaning a passenger and 0 meaning no passenger. The data set contains 46927855 records of 14728 taxis. To display the spatial distribution of trajectory points in the study area, evenly divide the study area into 100 by 100 grids, count the number of records in each grid, and draw Fig. 6.

### C. Data Analysis and Visualization

The entire dataset contains a large number of trajectories, which cannot be displayed in one picture, so just one taxi's trajectories are drawn. Fig. 7 shows 21 orders' trajectories for one taxi in a whole day. Due to trajectories having different lengths and too many trajectories are easy to overlap, it is unsuitable to be displayed in a figure with a constant scale. So, Fig. 7 is divided into four subfigures with different view ranges and scales. In Fig. 7, the solid line represents the trajectory that has not been optimized, and the dotted line represents the trajectory that the Kalman filter has optimized. Fig. 7(a), with the largest scale, shows some orders with long travel trajectories and long service duration, most of which occur late at night. Fig. 7(d), with the smallest scale, can more clearly compare the trajectories before and after optimization, which shows some orders in the city center.

### D. Results

In order to more intuitively compare the accuracy of the trajectory before and after optimization, the deviation between the trajectory point and the nearest road is calculated, and accumulate the deviations by each order. The results show that the average deviation between trajectory points and urban roads is reduced by 33.6%, which indicates that the trajectory data optimization method based on Doppler velocimetry proposed in this study can reduce the deviation between trajectories and urban road networks and improve the quality of trajectory data. Notably, the deviation between trajectory points and urban roads is related to the error level but cannot directly represent it. For example, in the trajectory of the 19th order in Fig. 7(d), there is a section of the trajectory with a significant deviation. Still, the deviation value of the entire order is insignificant due to the lack of positioning points in this trajectory section. Sometimes the optimized trajectory cannot reduce the deviation, such as the 7th order in Fig. 7(c). At the overall level, the proposed optimization method can be considered efficient.

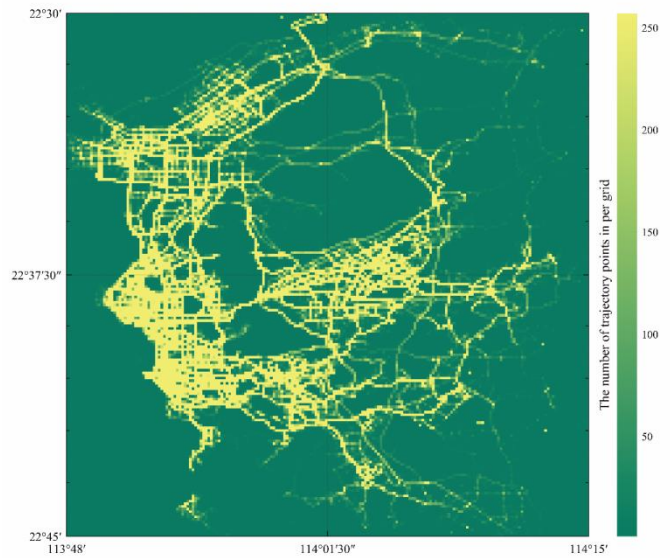


Fig. 6. The spatial distribution of trajectory points in the study area.

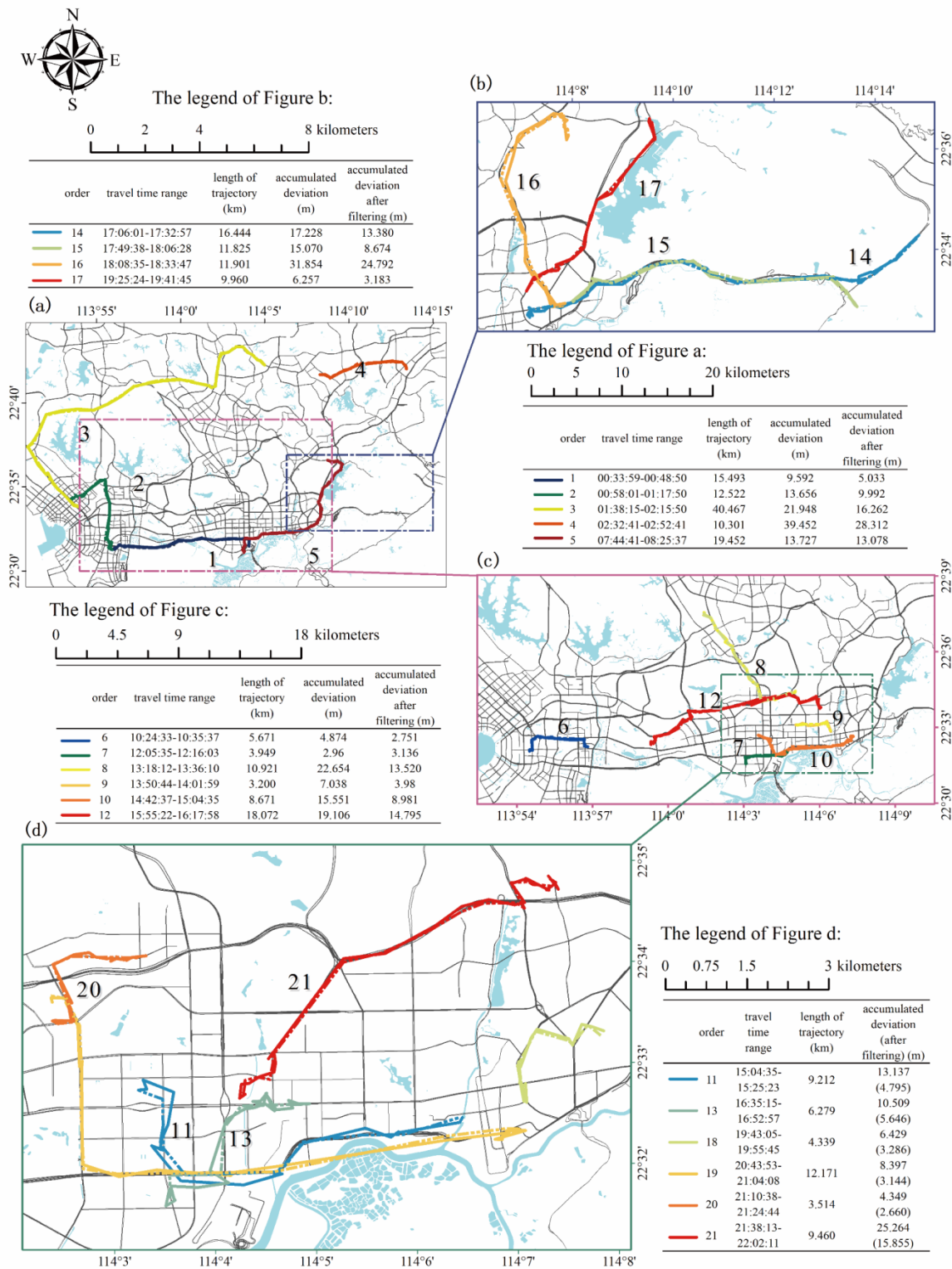


Fig. 7. Visualization of travel trajectories of a taxi before and after optimization.

### V. CONCLUSIONS

With the development of the mobile Internet, the number of terminal devices with positioning capability has increased rapidly. However, due to cost and energy consumption, the positioning capabilities of these devices have significant

differences. This paper establishes the system state equation of travel trajectory using Doppler velocity measurement and pseudo-range positioning data. By correlating the measured random error in the system state with the DOP of satellite positioning, a dynamic Kalman filter system is established to

optimize the trajectory data. The results show that the optimized trajectory data can better fit the urban road and can be applied to data analysis in transportation studies.

#### ACKNOWLEDGMENT

This research was funded by the National Natural Science Foundation of China No. 61963011, Guangxi Science and Technology Major Project No. AA19254016, Project of Natural Science Youth Foundation of Guangxi Province No. 2020JJB170049.

#### REFERENCES

- [1] Y. Zheng, "Trajectory Data Mining," *ACM Transactions on Intelligent Systems and Technology*. vol. 6, pp. 1-41, doi: 10.1145/2743025.
- [2] H. Feng, F. Bai and Y. Xu, "Identification of critical roads in urban transportation network based on GPS trajectory data," *Physica A: Statistical Mechanics and its Applications*. vol. 535, pp. doi: 10.1016/j.physa.2019.122337.
- [3] M. He, L. Zheng, W. Cao, et al., "An enhanced weight-based real-time map matching algorithm for complex urban networks," *Physica A: Statistical Mechanics and its Applications*. vol. 534, pp. doi: 10.1016/j.physa.2019.122318.
- [4] Y. Guo, D. Zou, X. Wang, et al., "Method for Estimating the Optimal Coefficient of LIC/BIC Signal Correlator Joint Receiving," *Remote Sensing*. vol. 14, pp. doi: 10.3390/rs14061401.
- [5] P. Steigenberger, S. Thoenert and O. Montenbruck, "GPS III Vespucci: Results of half a year in orbit," *Advances in Space Research*. vol. 66, pp. 2773-2785, doi: 10.1016/j.asr.2020.03.026.
- [6] W. Wang, Y. Wang, C. Yu, F. Xu and X. Dou, "Spaceborne atomic clock performance review of BDS-3 MEO satellites," *Measurement*. vol. 175, pp. doi: 10.1016/j.measurement.2021.109075.
- [7] J. A. Del Peral-Rosado, R. Raulefs, J. A. Lopez-Salcedo and G. Seco-Granados, "Survey of Cellular Mobile Radio Localization Methods: From 1G to 5G," *IEEE Communications Surveys & Tutorials*. vol. 20, pp. 1124-1148, doi: 10.1109/comst.2017.2785181.
- [8] A. Abdallah, J. Khalife and Z. M. Kassas, "Exploiting On-Demand 5G Downlink Signals for Opportunistic Navigation," *IEEE Signal Processing Letters*. vol. 30, pp. 389-393, doi: 10.1109/lsp.2023.3234496.
- [9] M. Pan, S. Liu, P. Liu, et al., "In Situ Calibration of Antenna Arrays for Positioning With 5G Networks," *IEEE Transactions on Microwave Theory and Techniques*. vol. pp. 1-14, doi: 10.1109/tmtt.2023.3256532.
- [10] M. Koivisto, M. Costa, J. Werner, et al., "Joint Device Positioning and Clock Synchronization in 5G Ultra-Dense Networks," *IEEE Transactions on Wireless Communications*. vol. 16, pp. 2866-2881, doi: 10.1109/twc.2017.2669963.
- [11] S. Fan, W. Ni, H. Tian, Z. Huang and R. Zeng, "Carrier Phase-Based Synchronization and High-Accuracy Positioning in 5G New Radio Cellular Networks," *IEEE Transactions on Communications*. vol. 70, pp. 564-577, doi: 10.1109/tcomm.2021.3119072.
- [12] Z. Fu, Z. Tian, Y. Xu and C. Qiao, "A Two-Step Clustering Approach to Extract Locations from Individual GPS Trajectory Data," *ISPRS International Journal of Geo-Information*. vol. 5, pp. doi: 10.3390/ijgi5100166.
- [13] X. Zhang, L. Lauber, H. Liu, et al., "Research on the method of travel area clustering of urban public transport based on Sage-Husa adaptive filter and improved DBSCAN algorithm," *PLoS One*. vol. 16, pp. e0259472, doi: 10.1371/journal.pone.0259472.
- [14] X. Liu, J. Guan, R. Jiang, S. S. Ge and B. Chen, "Finite-Horizon URTSS-Based Position Estimation for Urban Vehicle Localization," *IEEE Sensors Journal*. vol. 23, pp. 4011-4021, doi: 10.1109/jsen.2023.3235519.
- [15] H. Zhang, X. Xia, M. Nitsch and D. Abel, "Continuous-Time Factor Graph Optimization for Trajectory Smoothness of GNSS/INS Navigation in Temporarily GNSS-Denied Environments," *IEEE Robotics and Automation Letters*. vol. 7, pp. 9115-9122, doi: 10.1109/lra.2022.3189824.
- [16] D. Weaver Adams, C. Peck and M. Majji, "Doppler Light Detection and Ranging-Aided Inertial Navigation and Trajectory Recovery," *Journal of Guidance, Control, and Dynamics*. vol. pp. 1-19, doi: 10.2514/1.G007318.
- [17] G. Y. Li, Z. F. Huang, L. Y. Lou and P. J. Zheng, "Route Restoration Method for Sparse Taxi GPS trajectory based on Bayesian Network," *Teh Vjesn*. vol. 28, pp. 668-677, doi: 10.17559/Tv-20200513124207.
- [18] Z. Z. M. Kassas, M. Maaref, J. J. Morales, J. J. Khalife and K. Shamei, "Robust Vehicular Localization and Map Matching in Urban Environments Through IMU, GNSS, and Cellular Signals," *IEEE Intelligent Transportation Systems Magazine*. vol. 12, pp. 36-52, doi: 10.1109/imits.2020.2994110.
- [19] R. Santerre and A. Geiger, "Geometry of GPS relative positioning," *GPS Solutions*. vol. 22, pp. doi: 10.1007/s10291-018-0713-2.
- [20] S. Yaseen, F. Zafar and H. H. Alsulami, "An Efficient Jarratt-Type Iterative Method for Solving Nonlinear Global Positioning System Problems," *Axioms*. vol. 12, pp. doi: 10.3390/axioms12060562.
- [21] R. Santerre, A. Geiger and S. Banville, "Geometry of GPS dilution of precision: revisited," *GPS Solutions*. vol. 21, pp. 1747-1763, doi: 10.1007/s10291-017-0649-y.
- [22] J. Zhang, K. Zhang, R. Grenfell and R. Deakin, "On Real-Time High Precision Velocity Determination for Standalone GPS Users," *Survey Review*. vol. 40, pp. 366-378, doi: 10.1179/003962608x325420.
- [23] J. Feltens, G. Bellei, T. Springer, et al., "Tropospheric and ionospheric media calibrations based on global navigation satellite system observation data," *Journal of Space Weather and Space Climate*. vol. 8, pp. doi: 10.1051/swsc/2018016.
- [24] Y. Pei, S. Biswas, D. S. Fussell and K. Pingali, "An elementary introduction to Kalman filtering," *Communications of the ACM*. vol. 62, pp. 122-133, doi: 10.1145/3363294.
- [25] Y. Li, G. Chen and Y. Zhang, "Cycle-based signal timing with traffic flow prediction for dynamic environment," *Physica A: Statistical Mechanics and its Applications*. vol. 623, pp. doi: 10.1016/j.physa.2023.128877.
- [26] J. Zhou, T. X. Li, B. Chen and L. Yu, "Intermediate-variable-based Kalman filter for linear time-varying systems with unknown inputs," *International Journal of Robust and Nonlinear Control*. vol. 32, pp. 2453-2464, doi: 10.1002/rnc.5937

# Optimized YOLOv7 for Small Target Detection in Aerial Images Captured by Drone

Yanxin Liu, Shuai Chen\*, Lin Luo

School of Information and Control Engineering, Liaoning Petrochemical University, Fushun, China

**Abstract**—It is challenging to detect small targets in aerial images captured by drones due to variations in target sizes and occlusions arising from the surrounding environment. This study proposes an optimized object detection algorithm based on YOLOv7 to address the above-mentioned challenges. The proposed method comprises the design of a Genetic Kmeans (1-IoU) clustering algorithm to obtain customized anchor boxes that more significantly apply to the dataset. Moreover, the SPPFCSPC\_group structure is optimized using group convolutions to reduce model parameters. The fusion of Spatial Pyramid Pooling-Fast (SPPF) and Cross Stage Partial (CSP) structures leads to increased detection accuracy and enhanced multi-scale feature fusion network. Furthermore, a Detect Head is incorporated into the classification phase for more accurate position and class predictions. According to experimental findings, the optimized YOLOv7 algorithm performs quite well on the VisDrone2019 dataset in terms of detection accuracy. Compared with the original YOLOv7 algorithm, the optimized version shows a 0.18% increase in the Average Precision (AP), a reduction of 5.7 M model parameters, and a 1.12 Frames Per Second (FPS) improvement in the frame rate. With the above-described enhancements in AP and parameter reduction, the precision of small target detection and the real-time detection speed are increased notably. In general, the optimized YOLOv7 algorithm offers superior accuracy and real-time capability, thus making it well-suited for small target detection tasks in real-time drone aerial photography.

**Keywords**—Small target detection; drone aerial photography; YOLOv7; clustering algorithm; spatial pyramid pooling

## I. INTRODUCTION

Modern urban areas are characterized by dense city blocks, tall buildings, high population density, and heavy traffic, and they are capable of creating complex and dynamic environments. Satellite remote sensing is subjected to limitations in capturing high-resolution and high-dynamic range information for small targets for its revisit cycles, spatial resolution, and urban canyon effects. As sensor technology has been leaping forward, Unmanned Aerial Vehicles (UAVs) equipped with various sensors have emerged as effective tools for dynamically acquiring target images. UAV aerial imaging offers several advantages (e.g., a wide field of view, strong target detection capability, high real-time performance, as well as comprehensive information acquisition). Accurate detection and recognition of small targets through UAV aerial imaging enable fine-grained monitoring and provide valuable data for data-driven decision-making. However, conventional object detection algorithms struggle to effectively localize and accurately recognize small targets due to their low resolution and high noise interference.

Deep learning-based object detection algorithms have become the mainstream method due to their optimized efficiency and accuracy. The above-mentioned algorithms typically employ two-stage or one-stage detection strategies. Two-stage detection methods generate a series of candidate object boxes, which are subsequently filtered and refined by classifiers. Examples of two-stage algorithms include Faster Region-based Convolutional Neural Network (Faster R-CNN) [1] and Region-based Fully Convolutional Network (R-FCN) [2]. One-stage detection methods utilize convolutional neural networks [3] to extract image features and perform object classification and localization based on the above-described features. Algorithms such as You Only Look Once (YOLO) [4]–[11] and Single Shot MultiBox detector (SSD) [12] offer higher accuracy and generalization capability. To be specific, YOLOv7 has been confirmed as an advanced detection algorithm in the YOLO series, surpassing previous versions for inference speed and detection accuracy. Besides, it exhibits enhanced performance in detecting targets at different scales. However, challenges remain when YOLOv7 is employed for small target detection in UAV aerial imaging. First, small targets exhibit weak feature representation, such that they turn out to be susceptible to background interference and result in issues (e.g., false positives and false negatives). Second, deep learning models require significant computational resources for training and inference, whereas UAV aerial systems are subjected to limited computing resources and storage capacity. Accordingly, improving model size and computational efficiency becomes necessary. Lastly, deep learning algorithms are dependent on large-scale, high-quality annotated datasets to enhance their generalization capability, which is challenging to obtain specifically tailored for small target detection in UAV aerial imaging.

In this study, an enhanced YOLOv7 algorithm is presented for detecting small targets in UAV aerial imaging, to tackle the above challenges and fulfill the improvement requirements. The VisDrone2019 dataset is employed as the benchmark for detection. The proposed algorithm incorporates several significant enhancements, which comprise the redesign of anchor box sizes using an optimized clustering algorithm, the reduction of unnecessary candidate boxes, the reconstruction of the Spatial Pyramid Pooling (SPP) module, the integration of group convolutions and improved pooling connections, the reduction of model parameters, and the increased detection efficiency. Furthermore, a more accurate detection head, termed Detect, is introduced for target classification and position regression. The specific contributions of this study are elucidated below:

- The design of a high-precision anchor box clustering algorithm, termed Genetic Kmeans (1-IoU), employs genetic algorithms to optimize Kmeans clustering and adopts Intersection over Union (IoU) distance as a novel distance metric. The above-described algorithm leads to higher detection accuracy while reducing the likelihood of missing small targets.
- The optimization of the SPP module, SPPFCSPC\_group, by integrating group convolutions and combining the SPPF module and the CSP structures. This enhancement improves the ability exhibited by the algorithm to detect multi-scale targets and reduces model complexity while increasing object detection accuracy.
- The adoption of a more precise detection head, termed Detect, achieves higher precision and recall in target classification and localization. Accordingly, false positives are reduced significantly, and the model is endowed with the enhanced ability to distinguish between targets and the background.

The optimized YOLOv7 algorithm is assessed on 10 target categories. Comparative analysis with the baseline YOLOv7 model demonstrates a 0.18% increase in Average Precision (AP), a reduction of 5.7% in model parameters, and a 1.12 times improvement in Frames Per Second (FPS). As revealed by the experimental results, the optimized YOLOv7 algorithm achieves high precision and speed in the recognition of tiny objects during UAV aerial imagery.

## II. RELATED WORK

In object detection, small targets are commonly defined in accordance with the relative scale, with a bounding box area to image area ratio less than the square root of 0.33, or following the absolute scale, with a resolution less than 32 by 32 pixels. In UAV aerial imaging, tiny target detection requires adjustments in data format, algorithm structure, and parameter settings to tackle several challenges (e.g., small target size, weak feature representation, occlusion, deformation, high noise interference, and real-time requirements). In general, researchers address the above-mentioned challenges by implementing multi-scale detection strategies to cope with small targets of different sizes, incorporating contextual information and spatial constraints to increase the target localization accuracy, and introducing attention mechanisms to handle complex scenarios with occlusions and deformations involved.

For algorithm optimization, Zhang et al. [13] proposed YOLOv7-RAR algorithm for urban vehicle recognition. To be specific, these researchers reconstructed the backbone network using the Res3Unit structure, with the aim of enhancing the model's capability to capture more nonlinear features. Moreover, they introduced an ACmix attention mechanism to address weak target localization arising from background interference. Zhu et al. [14] developed TPH-YOLOv5 algorithm for target detection in UAV captured scenes. In the above-described method, YOLOv5 serves as the baseline model, a Transformer prediction head is employed, and a Convolutional Block Attention Module (CBAM) attention

mechanism is incorporated to enhance detection performance in dense aerial target scenarios. The enhanced algorithm achieves a 7% increase in accuracy compared with the baseline YOLOv5 model. However, the above-described methods often introduced additional network layers and parameters, resulting in increased computational complexity and limiting practical applications.

For data preprocessing, augmenting the training dataset can lead to the enhanced diversity and quantity of small targets, such that the model can be endowed with the enhanced generalization capability. Optimizing anchor box strategies can reduce computational costs and improve the matching between anchor boxes and real targets, enhancing detection accuracy. For instance, Liu and Wang [15] developed a YOLO-based detection network for corn detection and used a technique for data synthesis to create simulated images of broken maize from genuine corn photographs, such that the challenge of acquiring training data for damaged corn can be addressed. In the task of insulator defect detection, Zheng et al. [16] optimized YOLOv7 algorithm using the Kmeans++ clustering algorithm to cluster insulator targets and generate anchor boxes that more significantly apply to the detection of insulator defects. In the video surveillance vehicle detection task, Pan et al. [17] designed the improved YOLOv5s algorithm using Kmeans algorithm to correct the anchor frames and coordinated the CA attention mechanism for image recognition, which provided more accurate vehicle detection results and higher efficiency in terms of processing speed. The proposed method achieved high detection accuracy and speed on NVIDIA TX2 platform. However, optimized anchor boxes may struggle to accurately differentiate targets when they are occluded or overlapped, such that the detection accuracy can be reduced.

To conform to real-time requirements, algorithm optimization techniques (e.g., network pruning and quantization) are capable of reducing model computation and memory usage, such that the inference process can be expedited. Moreover, computational complexity can be reduced using lightweight model structures. Wu et al. [18] employed pruning techniques to lightweight the YOLOv4 network for concrete crack detection, where the EvoNorm-S0 algorithm was adopted to increase the detection accuracy. The resulting model achieved a high mAP value of 92.54% and a 15.9% reduction in the inference time, such that a real-time and high-precision detection algorithm was yielded. With the aim of detecting rice diseases and pests, Jia et al. [19] improved the YOLOv7 method and used the lightweight MobileNetV3 network for feature extraction, such that the model parameters were reduced, while an accuracy of 92.3% was generated. However, lightweight structures or network pruning may reduce model capacity while adversely affecting its representation capability, particularly in complex scene tasks, such that the accuracy and generalization capability can be decreased.

Despite the advancements achieved by regulating network structures, existing network architectures still struggle to reconcile detection speed and accuracy, particularly in highly overlapping small target areas. Thus, in-depth improvements should be made to increase the speed and precision of small target recognition algorithms based on UAV aerial imagery,

ultimately elevating the capabilities of small target recognition and fine-grained monitoring in UAV aerial photography.

### III. OPTIMIZED YOLOV7 ALGORITHM

#### A. Overview of YOLOv7

YOLOv7 refers to a single-stage deep learning-based object detection framework that achieves efficient object detection by detecting all objects in a single forward pass [11]. Compared with previous versions of the YOLO series, YOLOv7 offers faster convolution operations and higher detection accuracy, enabling it to detect more fine-grained objects while maintaining high detection speed. The YOLOv7 network structure comprises three components, i.e., a backbone network, a feature pyramid pooling layer, and a Head. Fig. 1 presents the simplified diagram of the YOLOv7 network structure. Backbone utilizes multiple convolutional layers to extract rich feature information, which is employed for subsequent object detection. The neck structure introduces the Path Aggregation-FPN (PAFPN) structure, combining feature maps at different scales to endow the algorithm with the enhanced capability to recognize various-sized things. The head layer employs the RepConv structure in conjunction with the IDetect Head to predict the target class and bounding boxes from the feature maps. The YOLOv7 algorithm exhibits high speed and real-time object detection capabilities while finding wide applications in areas (e.g., real-time video surveillance, UAV aerial imaging, and autonomous driving). It is capable of expediting the realization of intelligent and automated applications in a wide variety of scenarios.

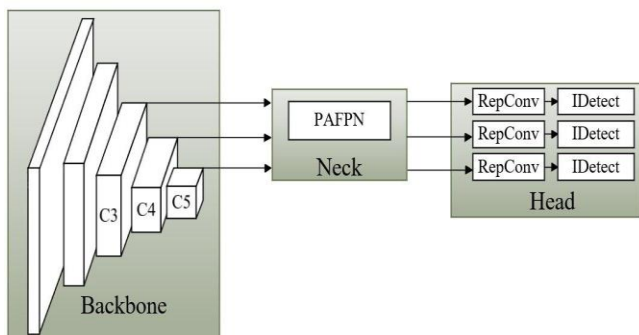


Fig. 1. Simplified diagram of YOLOv7 network structure.

However, YOLOv7 has high memory usage and may not be advantageous for mobile devices or resource-constrained systems. Additionally, the default anchor boxes of YOLOv7 are clustered based on the entire Common Objects in Context (COCO) training set, which may result in significant differences in target sizes and aspect ratios compared with the targets in specific detection scenarios. Accordingly, it is necessary to optimize and improve the YOLOv7 algorithm to better adapt to practical detection tasks and achieve superior detection performance.

#### B. Overall Structure of the Optimized YOLOv7 Network

In the optimized YOLOv7 object detection algorithm, YOLOv7 serves as the baseline model, and optimizations and improvements are introduced in three aspects (i.e., clustering anchor box sizes, SPP structure, and detection head). Fig. 2 presents the overall structure of the optimized YOLOv7. At the

preprocessing stage, the Genetic Kmeans (1-IoU) clustering algorithm is proposed in this study to redefine the shape of anchor boxes. The above-described algorithm adopts genetic algorithms to optimize Kmeans clustering while employing IoU distance as a distance metric. Based on this method, the redefined anchor box shapes are more significantly consistent with the custom sample data, such that the detection accuracy can be improved, and the false positives can be reduced. The spatial pyramid structure in the feature fusion network divides the feature map into various groups via group convolution, and each group is then subjected to convolution processes independently. Moreover, the SPPF module with a serial structure is combined with the CSP structure to decrease computational costs and increase the effectiveness of the receptive field. This combination forms the SPPFCSPC\_group module, which reduces the number of parameters, accelerates inference speed, and enhances the generalization ability of the model. The head layer incorporates RepConv module and Detect Head. By stacking multiple convolutional layers and sharing weights, the model can enhance its capacity to represent features and better comprehend the target's finer nuances. Moreover, RepConv module can adapt to targets of different scales and shapes. When combined with the Detect Head, it can be applied to feature maps at a wide range of levels, enhancing the model's capacity to recognize targets of all sizes and forms.

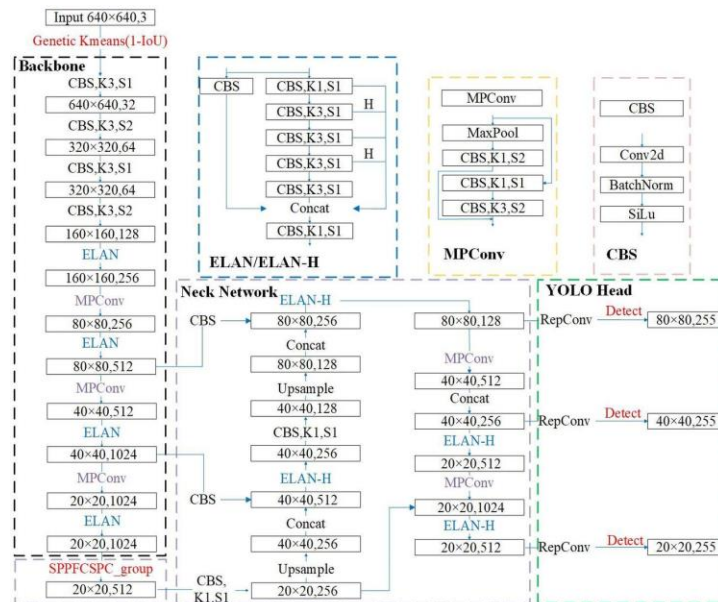


Fig. 2. Overall structure of optimized YOLOv7 network.

#### C. Genetic Kmeans (1-IoU) Anchor Box Clustering Algorithms

At the preprocessing stage of the object detection algorithm, this study proposes Genetic Kmeans (1-IoU) algorithm to recluster the anchor box shapes. Genetic Kmeans (1-IoU) algorithm utilizes genetic algorithms to optimize Kmeans clustering [20]. Following the random initialization of the population and iterative optimization through genetic operations, the problem of local optima is addressed, and clustering quality is improved [21]. Furthermore, under the presence of significant overlap between different scales and

categories in the dataset employed in this study, the conventional Kmeans algorithm employs Euclidean distance as the distance measure between sample points without considering the size and overlap of the object bounding boxes. This increases the uncertainty of the model regarding the object bounding boxes. Thus, Genetic Kmeans (1-IoU) algorithm introduces IoU distance, taking into account the separation between the center points and the overlap of the two bounding boxes. To be specific, this algorithm measures the similarity between different categories by calculating the IoU distance between the cluster centers and sample points.

The specific steps of Genetic Kmeans (1-IoU) algorithm are elucidated below:

1) Randomly select  $k$  samples as the initial centers of the clusters and randomly initialize the cluster centers. Determine the IoU distance between each sample's location and the center of each cluster, then place the sample in the cluster to which it is closest. The calculation equations are written in Eq. (1) and Eq. (2).

$$d(box, centroid) = 1 - IoU(box, centroid) \quad (1)$$

$$\mathcal{L}_{IoU} = 1 - IoU = 1 - \frac{B \cap B^{gt}}{B \cup B^{gt}} \quad (2)$$

2) Transform the clustering problem into an optimization problem of assessing the objective function, which can be written as:

$$\min J_E = \sum_{j=1}^c \sum_{k=1}^{n_j} \|x_k^j - m_j\| \quad (3)$$

$$\max J_B = \sum_{j=1}^c (m_j - m)^T (m_j - m) \quad (4)$$

where  $x_k^j$  represents the  $k$ -th sample that falls into the class;  $n_j$  denotes the number of samples in class  $j$ ;  $m_j$  expresses the center of class  $c_j$ , which is determined by Eq. (5);  $m$  represents the center of all samples, which is written in Eq. (6).

$$m_j = \frac{1}{n_j} \sum_{k=1}^{n_j} x_k^j \quad (5)$$

$$m = \frac{1}{n} \sum_{i=1}^n x_i \quad (6)$$

3) The clustering performance of the genetic algorithm is assessed using the *fitness* value, as shown in Eq. (7). A higher fitness value indicates a greater likelihood for the individual's genes to be selected for the next generation.

$$fitness = \frac{J_B}{J_E} = \frac{\sum_{j=1}^c (m_j - m)^T (m_j - m)}{\sum_{j=1}^c \sum_{k=1}^{n_j} \|x_k^j - m_j\|} \quad (7)$$

#### D. SPPFCSPC\_Group Structure

Based on the SPPCSPC module (as shown in Fig. 3(a)), the SPPFCSPC\_group module is designed to perform feature fusion and dimensionality reduction at different scales in the Neck network structure of the improved YOLOv7 algorithm. Fig. 3(b) presents the structure of the SPPFCSPC\_group module, comprising a series of group convolutions, SPPF module, and CSP structure. By partially connecting at different stages of the network and cross-linking the features of the earlier stage with the later stage, the SPPFCSPC\_group module

increases the performance of target recognition and the network's capacity to represent features.

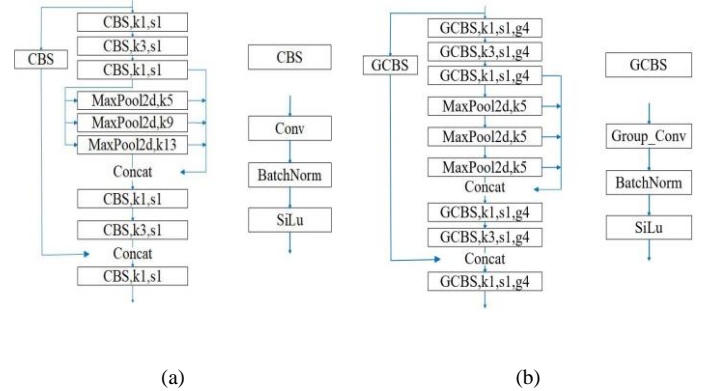


Fig. 3. Space pyramid pooling module (a) Structure of SPPCSPC module (b) Structure of SPPFCSPC\_group module.

To be specific, the input image undergoes feature extraction through a series of group convolution layers. After feature extraction, the SPPF module uses group convolutions to execute multi-scale pooling operations to capture broad and specific information at various scales. The input feature map is divided into various scales by the SPPF module, and each scale undergoes a group convolution operation to obtain scale-specific feature representations [22]. The group convolutions concatenate the feature maps from multiple scales, resulting in a feature representation that contains global and local information at different scales. After feature fusion, The CSP module separates the feature map into two parts after feature fusion: one portion directly conducts the subsequent convolution operation; the other half is preprocessed before being fused with the previous component, such that the feature representation capability can be enhanced.

Fig. 4 depicts the structure of group convolution. Eq (8) and (9), respectively, indicate the number of parameters in a single convolution kernel and the total number of parameters in the convolution layer.

$$P_1 = \begin{cases} C_{in} \times K_1 \times K_2, bias = False \\ C_{in} \times K_1 \times K_2 + 1, bias = True \end{cases} \quad (8)$$

$$Parameters = C_{out} \times P_1 \quad (9)$$

where the input is expressed as  $C_{in}$ ,  $H_{in}$ ,  $W_{in}$ , and  $D_{in}$ ; the output is denoted as  $C_{out}$ ,  $H_{out}$ ,  $W_{out}$ , and  $D_{out}$ .

Group convolution divides the input feature map into  $g$  groups following the channel dimension while applying a regular convolution to the respective group. The number of parameters for group convolution is represented by Eq. (10).

$$Parameters_{Group} = \begin{cases} C_{out} \times (\frac{C_{in}}{g} \times K_1 \times K_2), bias = False \\ C_{out} \times (\frac{C_{in}}{g} \times K_1 \times K_2 + 1), bias = True \end{cases} \quad (10)$$

where  $\frac{C_{in}}{g}$  denotes the number of channels in each group of the input feature map, i.e., the number of channels in the respective convolutional kernel. After group convolution is

completed, a regular convolution is applied to the respective group. Since the respective group requires at least one convolutional kernel, the output channel number  $C_{out}$  for group convolution is at least  $g$ . If the respective group covers  $n$  convolutional kernels, the output channel number  $C_{out}$  is given by  $n \times g$  ( $n > 1$ ), here  $y$  expresses the number of groups. In other words, the output channel number  $C_{out}$  is a multiple of the number of groups. Accordingly, group convolution requires that the input and output channel numbers be evenly divided by the number of groups  $g$ . The reduction in the number of parameters is the fundamental reason behind the decrease in the channel number of the respective convolutional kernel to  $1/g$  after group convolution is completed.

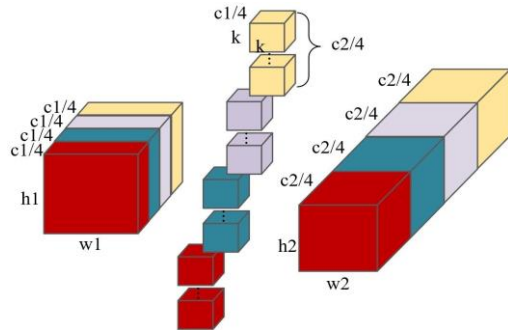


Fig. 4. Structure of group convolution.

### E. Optimized YOLOv7 Detection Head

The optimized YOLOv7 algorithm utilizes the Detect Head in the Head layer to obtain prediction results. The feature map transfers the Detect Head by the optimized YOLOv7 model. Fig. 5 depicts the Detect module's flowchart. The Detect Head utilizes a series of convolutional layers and fully connected layers to predict the position and class of the target. The above-described layers extract features through convolution operations and nonlinear activation functions and map them to the spatial coordinates and class of the target. The optimized YOLOv7 algorithm outputs the prediction results through the Detect Head, along with labels and confidence scores for the target classes. Since no extra classifiers or regressors are necessary because the Detect Head can anticipate the bounding boxes and classes of the targets directly, the model structure can be simplified, computational and memory overhead can be reduced, and the inference speed can be increased. Furthermore, the Detect Head is not dependent on feature vectors to predict the position and size of the targets, such that the bounding boxes and classes can be directly predicted. Consequently, the Detect Head enhances the precision of target localization to a certain extent.

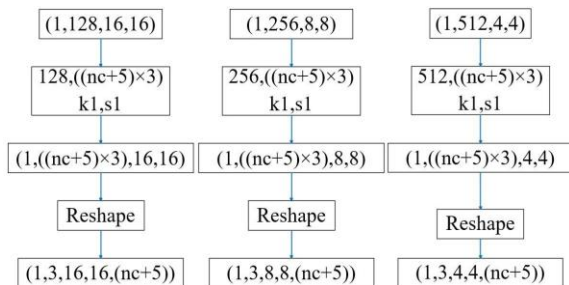


Fig. 5. Flowchart of detection head.

## IV. EXPERIMENTAL VALIDATION AND ANALYSIS

### A. Dataset Preparation

In this study, the performance of the optimized YOLOv7 model for small object detection is assessed using the VisDrone2019 dataset. A total of 2158 samples are randomly selected to generate a custom dataset, with the aim of investigating the detection capabilities of the optimized YOLOv7 algorithm on small objects. A training set and a test set are divided into the custom dataset in 7:3 ratio. The original detection categories are further assigned to 10 classes. For simplicity, new names are assigned to the above-mentioned 10 classes in the experiment. The names and distribution of the target categories are listed in Table I. In the custom dataset, based on the definition of small objects for relative scale, small objects account for approximately 70% of the dataset. Likewise, small objects take up approximately 54% of the dataset, following the definition of small objects for absolute scale.

TABLE I. CORRESPONDING NAMES AND QUANTITY DISTRIBUTION OF TARGET CATEGORIES

Category	Models	Accuracy (%)
pedestrian	C1	79339
people	C2	27059
bicycle	C3	10480
car	C4	144867
van	C5	24956
truck	C6	12875
tricycle	C7	4812
awning-tricycle	C8	3246
bus	C9	5926
motor	C10	29647

### B. Experimental Condition and Assessment Metrics

The experiments are performed on an Alienware X15 R1 laptop with the following hardware specifications: 11th Gen Intel (R) Core (TM) i7-11800H CPU (2.3GHz), 32GB RAM, NVIDIA GeForce RTX 3070 GPU with 8GB VRAM. The experiments are performed using the PyTorch deep learning framework on Windows 11 operating system. The program code is implemented in Python, utilizing libraries (e.g., CUDA, Cudnn, and OpenCV). The above-mentioned setup contributes to the training and testing of tiny item detection on the VisDrone2019 dataset. In the comparative experiments and fusion studies, the input image is configured to be 640 by 640 pixels. 50 total epochs of training are completed, with a batch size of 2. Weight decay is set to 0.0005, momentum is set to 0.937, and the initial learning rate is set to 0.01.

Common assessment metrics in object detection algorithms are employed to objectively evaluate the effectiveness of the detection models. The above-described metrics comprise AP, mean Average Precision (mAP), Number of Parameters (Params), Giga Floating-point Operations Per Second (GFLOPS), as well as FPS.



C. Comparative Analysis of Experimental Results

1) Comparison and analysis of clustering algorithm loss curves: During the training of the YOLOv7 model, the Best Possible Recall (BPR) between the default anchor boxes and each target in the custom dataset is automatically determined by the network. If the BPR falls below 0.98, the detection model uses a combination of genetic algorithm and Kmeans to recluster and generate new anchor boxes, known as Autoanchor. Autoanchor combines genetic algorithm with Kmeans clustering and utilizes Euclidean distance for mutation on the clustering results. The experiment tested three different clustering algorithms: Autoanchor using Euclidean distance, Kmeans using 1-IoU distance, and Genetic Kmeans. Table II displays the predicted anchor box forms for each clustering algorithm at various scales.

TABLE II. ANCHOR BOX SHAPES FOR THREE CLUSTERING ALGORITHMS AT PREDICTED SCALES

Branch	P3	P4	P5
Dimension	80 × 80	40 × 40	20 × 20
Autoanchor	(2,3,3,8,6,5)	(7,14,12,8,21,12)	(13,21,30,23,48,45)
Kmeans (1-IoU)	(2,3,3,7,6,5)	(6,12,11,9,11,20)	(22,12,25,25,47,39)
Genetic Kmeans (1-IoU)	(2,4,3,8,6,6)	(6,12,12,9,11,18)	(25,14,23,28,47,36)

Table III presents a comparison of assessment metrics for a variety of clustering algorithms. The models with optimized anchor box sizes achieve overall improved detection accuracy compared with the baseline network. mAP achieved by the model using Genetic Kmeans (1-IoU) as the clustering algorithm reaches 31.8%, marking improvements of 0.4% and 0.4% compared with Autoanchor and Kmeans (1-IoU), respectively. To be specific, mAP is raised by 0.9% in comparison to the original YOLOv7 algorithm. Furthermore, AP obtained by training the network with Genetic Kmeans (1-IoU) reaches 17.04%, marking improvements of 1.41% and 0.61% over the original YOLOv7 algorithm and Autoanchor, respectively. As revealed by the above-mentioned results, the detection model achieves higher detection accuracy by using 1-IoU distance and improving Kmeans clustering method with genetic algorithms to generate anchor boxes that more effectively match the sizes of detection targets in the sample dataset. In general, compared with the original YOLOv7 model, Genetic Kmeans (1-IoU) achieves the optimal performance.

TABLE III. COMPARISON OF ASSESSMENT METRICS FOR DIFFERENT CLUSTERING ALGORITHMS

Anchor	YOLOv7	Autoanchor	Kmeans (1-IoU)	Genetic Kmeans (1-IoU)
AP (%)	15.63	16.43	17.05	17.04
mAP (%)	30.9	31.4	31.4	31.8
GFLOPS	103.5	103.5	103.5	103.5
FPS	60.61	65.79	65.79	64.93
Params (M)	36.54	36.54	36.54	36.54

The trained models are further validated for loss. Fig. 6 presents the comparison of loss curves for different clustering algorithms. As depicted in Fig. 6, Autoanchor and Kmeans (1-IoU) have slightly higher losses compared with Genetic Kmeans (1-IoU), with average losses of 0.13, 0.1309, and 0.1288, respectively. In contrast, YOLOv7 has the slowest decrease in loss, with a final average loss of 0.1316. The above result suggests that the Genetic Kmeans (1-IoU) algorithm reduces the loss of the original YOLOv7 algorithm by 0.28%.

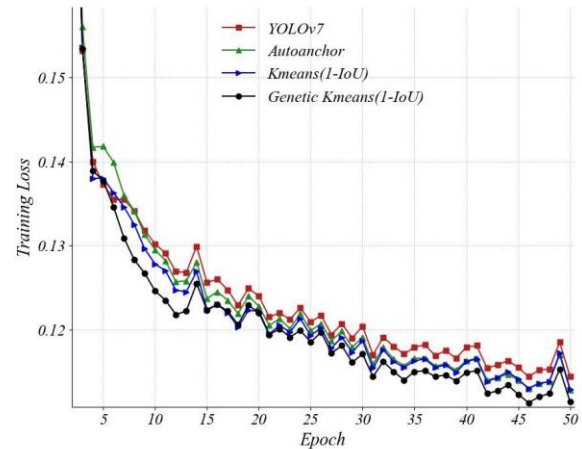


Fig. 6. Comparison of loss curves for different clustering algorithms.

2) Comparison and analysis of assessment metrics for pyramid pooling structure: A fusion experiment is performed on the pyramid pooling module to validate the effectiveness of the SPPFCSPC\_group module, which utilizes grouped convolution and the SPPF structure, in small object detection from aerial images captured by drones. Starting with the SPPCSPC module, improvements are made sequentially with grouped convolution and the SPPF structure. Table IV presents the comparison of assessment metrics for the fusion study of the pyramid pooling module. As seen in Table IV, using grouped convolution decreases the module’s parameter size by 5.7 M. The SPPFCSPC\_group structure achieves an AP value of 15.82%, marking an improvement of 0.19% compared with the SPPCSPC structure. Moreover, FPS is increased by 3.91, validating the performance of the optimized structure for accuracy and speed.

TABLE IV. PERFORMANCE COMPARISON OF ASSESSMENT METRICS IN FUSION STUDY ON PYRAMID POOLING MODULE

Neck	SPPCSPC	SPPCSPC_group	SPPFCSPC_group
AP (%)	15.63	15.26	15.82
mAP (%)	30.9	30.8	30.8
GFLOPS	103.5	99.0	99.0
FPS	60.61	60.98	64.52
Params (M)	36.54	30.84	30.84

To further verify the effect of the SPPFCSPC\_group module on the detection accuracy of small object samples in the YOLOv7 model for aerial drone images, experiments are performed to compare five different pyramid pooling structures, i.e., SPP [23], SPPF, Atrous Spatial Pyramid Pooling (ASPP) [24], Receptive Field Block (RFB) [25], and SPPFCSPC\_group. Table V presents the comparison of assessment metrics for the comparative study of the pyramid pooling module.

As depicted in Table V, the YOLOv7 baseline network achieves the maximum AP value and mAP value when using the SPPFCSPC\_group structure, which is 15.82% and 30.8% respectively. The adoption of group convolution reduces the model parameters to only 30.84 M, resulting in a reduction of 14.61 and 2.44 M compared with the ASPP module and the RFB module that utilize dilated convolution, respectively. As revealed by the above result, while reducing the model parameters, the model maintains a high detection accuracy and avoids weakening the information interaction between different layers by using group convolution. The above-described findings validate the effectiveness of SPPFCSPC\_group module.

TABLE V. PERFORMANCE COMPARISON OF ASSESSMENT METRICS IN COMPARATIVE STUDY ON PYRAMID POOLING MODULE

Neck	SPP	SPPF	ASPP	RFB	SPPFCSPC_group
AP (%)	15.26	15.50	15.73	15.78	15.82
mAP (%)	30.6	30.7	30.6	30.5	30.8
GFLOPS	98.7	98.7	110.6	100.9	99.0
FPS	65.79	68.49	64.51	64.94	64.52
Params (M)	30.51	30.51	45.45	33.28	30.84

3) Comparison and analysis of different detection heads:

The positive and negative sample allocation strategy of YOLOv7 is designed around the Lead head and the Auxiliary head, combining the positive and negative sample allocation strategies of YOLOv5 and YOLOX. To assess the impact of different detection heads on the model’s detection accuracy, a comparative analysis was conducted among YOLOv5, YOLOX, the default Detect Head used in YOLOv7, Decoupled Head, and IDetect Head.

TABLE VI. PERFORMANCE COMPARISON OF DIFFERENT DETECTION HEADS IN DETECTORS

Head	IDetect	Decoupled Head	Detect Head
AP (%)	15.63	16.80	15.73
mAP (%)	30.9	24.7	30.4
GFLOPS	103.5	144.6	103.5
FPS	60.61	54.05	62.50
Params (M)	36.54	44.03	36.54

As depicted in Table VI, the Decoupled Head achieves the minimum overall precision, with a mAP value of only 24.7%, which is significantly lower than the IDetect Head (6.2%) and

the Detect Head (5.7%). Besides, the Decoupled Head also has the largest parameter count, exceeding that of IDetect by 7.49 M and Detect by 7.49 M. Furthermore, when using the Detect Head, the overall AP value reaches the maximum point at 15.73%, representing a 0.1% increase compared with IDetect Head, while keeping the model size unchanged.

In order to further observe the effect of different detection heads on the model detection accuracy, the experiments are shown in Fig. 7 as scatter plots of the P-R curves on the VisDrone dataset for detectors using different detection heads. The precision P is represented by the vertical axis, while the recall rate R is represented by the horizontal axis. The area enclosed by the curve and the coordinate axes represents the AP value, where a curve closer to the top right corner indicates a better detection model. As seen in Fig. 7, the recall rate steadily raises but the accuracy declines as the number of epochs rises. Decoupled Head shows a rapid decline in precision when the recall rate reaches 20%, suggesting a lower performance of the detector. The P-R curve of Detect largely envelops the curve of IDetect, demonstrating higher precision and recall. This indicates that Detect has stronger adaptability to different scenes, lighting conditions, and variations in target morphology. Thus, through experimental analysis and comparison, Detect achieves optimal classification performance, making it the preferred detection head for the detection model.

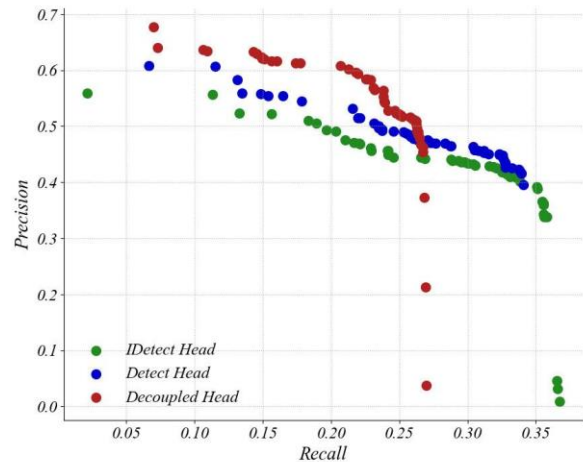


Fig. 7. P-R scatter plot of detectors with different detection heads on the VisDrone dataset.

4) Comparative analysis of fusion studies: To validate the effect of the suggested improvements on the detection model, fusion studies were conducted by testing the components of the improvement method. Table VII compares the results of the fusion studies. The fusion experiments were performed based on the YOLOv7 baseline model, and the improvements were incrementally added to observe their effects on the research objectives and assess their importance. First, IDetect Head was replaced with the Detect Head. Then, Genetic Kmeans (1-IoU) clustering algorithm was added. Finally, the

SPPFCSPC\_group module was added on top of the previous modifications.

TABLE VII. COMPARATIVE RESULTS OF FUSION STUDIES

Method	Baseline	+Detect Head	+Genetic Kmeans (1-IoU)	+SPPFCSPC_group
AP (%)	15.63	15.73	15.13	15.81
mAP (%)	30.9	30.4	30.0	30.3
GFLOPS	103.5	103.5	103.5	99.0
FPS	60.61	62.50	59.52	61.73
Params (M)	36.54	36.54	36.54	30.84

Table VII shows that the proposed improvements have achieved performance gains in small object detection from aerial images. First, replacing the Detect Head resulted in higher detection accuracy with a 0.1% increase in AP and a 2.4 FPS improvement in detection speed, suggesting the good performance of the Detect Head in the context of this study. Second, when the Genetic Kmeans (1-IoU) algorithm and SPPFCSPC\_group module were added on top of the Detect-based model, AP reached its maximum value at 15.81%, which is an improvement of 0.8% compared with YOLOv7. Additionally, the model's parameter count decreased to 30.84 M, which is a reduction of 5.7 M compared with YOLOv7, while achieving an FPS of 61.73, which is a 1.12 improvement over YOLOv7. The above-mentioned results demonstrate that the model may concentrate on positive anchor boxes of high quality by upgrading the original anchor boxes leading to increased detection accuracy. Moreover, the improvements in the form of group convolution and the SPPF module effectively reduce the model's parameter count while increasing the inference speed, thus conforming to the requirements for real-time detection.

5) Comparison and analysis of different models: Table VIII lists the comparative experimental results of a variety of algorithms. Two lightweight models (i.e., YOLOv5s and YOLOv7-Tiny) are tested and compared through the experiments. As depicted in Table VIII, YOLOv5s achieves a higher AP value than YOLOv7-Tiny by 1.45%, whereas its mAP value is 70.7% lower than that of YOLOv7-Tiny. As indicated by the above results, YOLOv5s exhibits high performance in certain categories, while YOLOv7-Tiny exhibits overall higher performance. Among other larger models, the YOLOR-P6 algorithm achieves the minimum detection accuracy, with an AP value of only 10.21% and an mAP value of 20.9%. Optimized YOLOv7 achieves the maximum AP value, with improvements of 0.89%, 3.63%, 5.6%, and 0.18% compared with YOLOv3-SPP, YOLOv51, YOLOR-P6, and YOLOv7, respectively. mAP value of Optimized YOLOv7 is 30.3%, with improvements of 1.0%, 4.8%, and 9.4% compared with YOLOv3-SPP, YOLOv51, and YOLOR-P6, respectively.

TABLE VIII. COMPARATIVE EXPERIMENTAL RESULTS OF DIFFERENT ALGORITHMS

Models	AP (%)	mAP (%)	GFLOPS	FPS	Params (M)
YOLOv3-SPP	14.92	29.3	155.7	54.05	62.61
YOLOv51	12.18	25.5	114.3	59.52	44.66
YOLOv5s	10.20	16.6	16.4	100.00	7.08
YOLOR-P6	10.21	20.9	80.4	63.29	36.87
YOLOv7-Tiny	8.75	17.3	13.2	70.42	6.04
YOLOv7	15.63	30.9	103.5	60.61	36.54
Optimized YOLOv7	15.81	30.3	99.0	61.73	30.84

To provide a more intuitive comparison of different models in detecting the same samples, the comparison of AP values for the respective category is presented in Fig. 8. For some categories with fewer samples, difficult discrimination, and occlusion, such as 'people', 'truck', and 'bicycle', the improvement in accuracy and speed of the models is limited. However, in general, there is a balance between the reduction in model parameters and the improvement in accuracy.

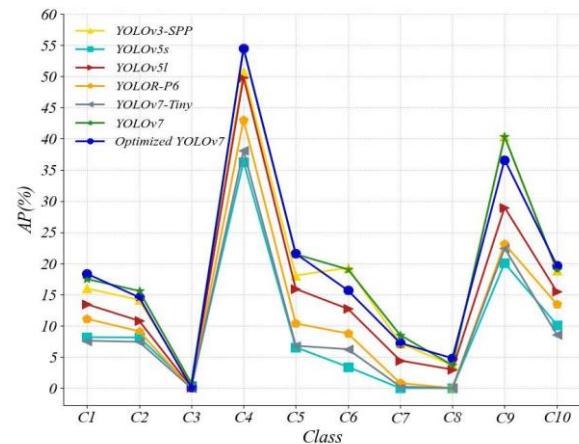


Fig. 8. Comparison of AP values for the respective category.

The detection results of YOLOv3-SPP, YOLOv51, YOLOR-P6, and the optimized YOLOv7 algorithm are compared for three different scenarios: slight category differences, dense object distribution, and low-light conditions at night. As depicted in Fig. 9, detection algorithms are more prone to false positives and false negatives when there exists slight category differences and dense object distribution. Under low-light conditions at night, the visibility of small objects declines significantly, such that the blurred details and edges are generated, adversely affecting the effective feature extraction of the detection network. In contrast to other algorithms, the optimized YOLOv7 algorithm is effective in mitigating the above described interference factors and demonstrates outstanding detection performance in various scenarios.

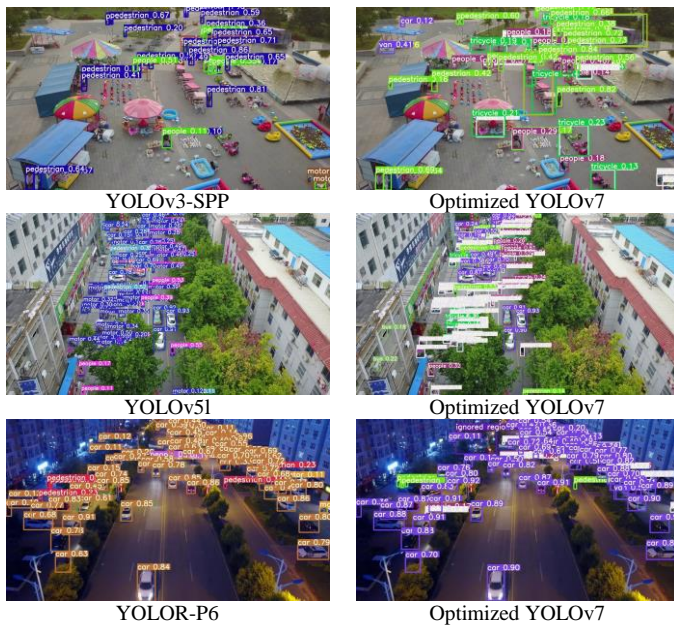


Fig. 9. Comparison of test outcomes among various algorithms.

6) *GradCAM heatmap visualization analysis*: In this study, the heatmaps of the 102nd layer of the detection network are visualized using GradCAM. The highlighted regions in the heatmaps represent the areas that the network considers relevant to the target categories. The 102nd layer represents the P3 branch of the model, i.e., the feature layer specifically developed for small object detection. This visualization presents a more intuitive insight into the network’s attention and decision-making process in detecting small objects. In the VisDrone dataset, categories (e.g., ‘pedestrian’, ‘people’, and ‘motor’) are considered small objects for their relatively small sizes. Moreover, ‘car’ can still be considered a small object category in scenarios with long distances and significant occlusions even it exhibits a larger relative size. During the experiment, feature visualization is performed for the above-mentioned four small object categories, and Fig. 10 presents the visualization of the detection image heatmaps.

As depicted in Fig. 10, the highlighted regions in the heatmaps represent the detected object positions, suggesting that the network can clearly concentrate on small targets. Furthermore, the intensity of colors in the heatmaps represents the degree of network attention. Compared with the YOLOv7 algorithm, the optimized YOLOv7 algorithm exhibits stronger intensity in the highlighted regions (① and ②) when detecting ‘pedestrian’ and ‘people’, suggesting that the optimized YOLOv7 algorithm accurately focuses on the target objects while exhibiting a higher level of attention towards small targets. In the visualization image for the ‘motor’ category, as indicated by the label (③), when a significant overlap exists between ‘people’ and ‘motor’, the YOLOv7 algorithm tends to produce false positives. However, the optimized YOLOv7 algorithm displays a more distinctive and accurate highlighting in the area representing the ‘motor’ target, suggesting improved attention towards the detection targets. For the ‘car’

category, under a long distance or tree occlusion, the attention intensity of YOLOv7 turns out to be weaker, such that potentially missed detections are conducted. In contrast, the optimized YOLOv7 algorithm achieves the notably enhanced color intensity in the heatmap, marked as (④), suggesting an improvement in detecting small objects that are previously missed.

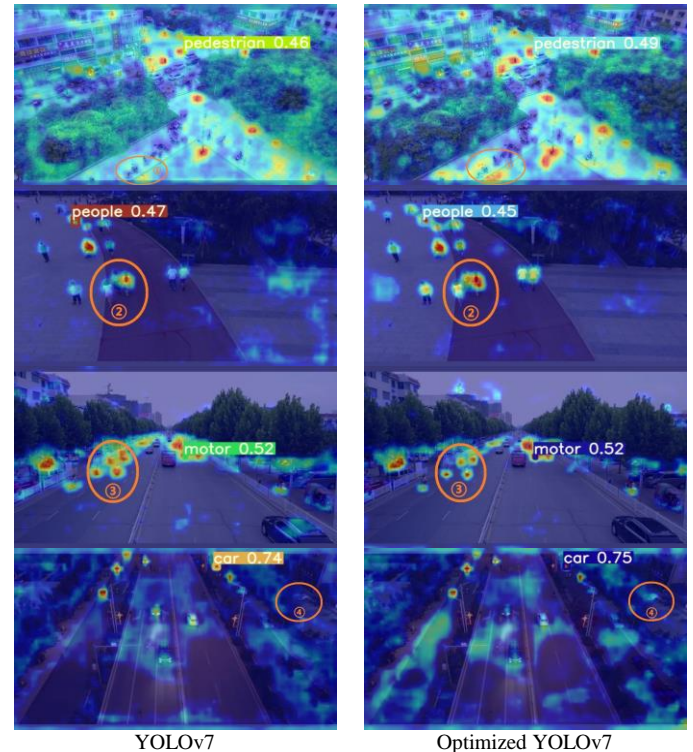


Fig. 10. Visualization of detection image heatmaps.

As revealed by the experimental analysis, the optimized YOLOv7 algorithm demonstrates significant advantages in accuracy and speed for locating tiny objects in aerial photographs taken by UAVs. In the comparative experiments, the proposed Genetic Kmeans (1-IoU) clustering algorithm allows the model to more effectively cluster the anchor box sizes for small targets. Moreover, the optimized SPPFCSPC\_group module, utilizing group convolution, effectively reduces the model parameters. The integration of the SPPF module with the CSP structure enhances both the speed of inference and the precision of detection. Lastly, the use of the Detect Head improves the model’s confidence in target detection. The optimized YOLOv7 algorithm is capable of recognizing small-sized objects in UAV aerial images more significantly, even in sophisticated backgrounds. Furthermore, fusion experiments are used to confirm the effectiveness of the proposed methods.

## V. CONCLUSION

In this study, an optimized YOLOv7 algorithm is proposed to address the challenges of detecting small-sized and heavily occluded objects in aerial images captured by UAVs. The proposed method comprises three key steps. At the preprocessing stage, an anchor box clustering algorithm is designed to achieve anchor boxes that better suit the dataset,

increasing the accuracy of object detection and reducing the rate of missed detections for small targets. In the feature fusion network, SPP structure based on group convolution is introduced to reduce model parameters and computational complexity. The inference speed of the model is enhanced by adopting a serial pyramid pooling method. Lastly, a detection head that is more tailored to the custom dataset is employed to refine the detection layers. With this method, more accurate detection of small-sized and low-count categories of objects can be achieved. Experimental findings show that compared with the standard YOLOv7, the suggested approach achieves an AP improvement of 0.18%, reduces the model size by 4.5 GFLOPS, decreases the network parameter size by 5.7 million, and increases FPS by 1.12. Accordingly, the proposed method enhances the applicability of the YOLO algorithm for locating tiny targets in aerial photographs that UAVs have recorded.

#### ACKNOWLEDGMENT

This work was supported by the Scientific Research Project of the Liaoning Provincial Department of Education (No.LJKZ0398, No.LJKZ0423) and the Scientific Research Foundation of Liaoning Petrochemical University (No.2017XJJ-044, No.2017XJJ-012). The authors also would thank reviewers for improving this article.

#### REFERENCES

- [1] 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.
- [2] Jifeng Dai, Yi Li, Kaiming He, and Jian Sun. R-fcn: Object detection via region-based fully convolutional networks. *Advances in neural information processing systems*, 29:1–9, 2016.
- [3] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.
- [4] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.
- [5] Joseph Redmon and Ali Farhadi. Yolo9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7263–7271, 2017.
- [6] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *CoRR*, abs/1804.02767, 2018.
- [7] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. Yolov4: Optimal speed and accuracy of object detection. *CoRR*, abs/2004.10934, 2020.
- [8] Chuyi Li, Lulu Li, Hongliang Jiang, Kaiheng Weng, Yifei Geng, Liang Li, Zaidan Ke, Qingyuan Li, Meng Cheng, Weiqiang Nie, Yiduo Li, Bo Zhang, Yufei Liang, Linyuan Zhou, Xiaoming Xu, Xiangxiang Chu, Xiaoming Wei, and Xiaolin Wei. Yolov6: A single-stage object detection framework for industrial applications, 2022.
- [9] Zheng Ge, Songtao Liu, Feng Wang, Zeming Li, and Jian Sun. YOLOX: exceeding YOLO series in 2021. *CoRR*, abs/2107.08430, 2021.
- [10] Chien-Yao Wang, I-Hau Yeh, and Hong-Yuan Mark Liao. You only learn one representation: Unified network for multiple tasks. *CoRR*, abs/2105.04206, 2021.
- [11] Chien Yao Wang, Alexey Bochkovskiy, and Hong Yuan Mark Liao. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7464–7475, 2023.
- [12] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng Yang Fu, and Alexander 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*, pages 21–37, 2016.
- [13] Yuan Zhang, Youpeng Sun, Zheng Wang, and Ying Jiang. Yolov7-rar for urban vehicle detection. *Sensors*, 23(4), 2023.
- [14] Xinghui Zhu, Shuchang Lyu, Xu Wang, and Qi Zhao. Tpholov5: Improved yolov5 based on transformer prediction head for object detection on drone-captured scenarios. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, pages 2778–2788, October 2021.
- [15] Zechuan Liu and Song Wang. Broken corn detection based on an adjusted yolo with focal loss. *IEEE Access*, 7:68281–68289, 2019.
- [16] Jianfeng Zheng, Hang Wu, Han Zhang, Zhaoqi Wang, and Weiyue Xu. Insulator-defect detection algorithm based on improved yolov7. *Sensors*, 22(22), 2022.
- [17] Yi Pan, Zhao Zhu, Yan Hu, and Qing Wang. Video surveillance vehicle detection method incorporating attention mechanism and yolov5. *International Journal of Advanced Computer Science and Applications*, 14(6), 2023.
- [18] Peirong Wu, Airong Liu, Jiyang Fu, Xijun Ye, and Yinghao Zhao. Autonomous surface crack identification of concrete structures based on an improved one-stage object detection algorithm. *Engineering Structures*, 272:114962, 2022.
- [19] Liangquan Jia, Tao Wang, Yi Chen, Ying Zang, Xiangge Li, Haojie Shi, and Lu Gao. Mobilenet-ca-yolo: An improved yolov7 based on the mobilenetv3 and attention mechanism for rice pests and diseases detection. *Agriculture*, 13(7), 2023.
- [20] Ahmed M, Seraj R, Islam S M S. The k-means algorithm: A comprehensive survey and performance evaluation[J]. *Electronics*, 2020, 9(8): 1295.
- [21] Lambora A, Gupta K, Chopra K. Genetic algorithm-A literature review[C]//2019 international conference on machine learning, big data, cloud and parallel computing (COMITCon). IEEE, 2019: 380-384. Yuxin Wu and Kaiming He. Group normalization. *CoRR*, abs/1803.08494, 2018.
- [22] Wu Y, He K. Group normalization[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.
- [23] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(9):1904–1916, 2015.
- [24] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *CoRR*, abs/1706.05587, 2017.
- [25] Songtao Liu, Di Huang, and Yunhong Wang. Receptive field block net for accurate and fast object detection. *CoRR*, abs/1711.07767, 2017.

# DevOps Implementation Challenges in the Indonesian Public Health Organization

Muhammad Yazid Al Qahar, Teguh Raharjo

Faculty of Computer Science, University of Indonesia, Jakarta, Indonesia

**Abstract**—The importance of accelerating software development to meet rapidly changing business needs has driven the Indonesian Public Health Organization (IPHO) to adopt DevOps. But after three years, the expected benefits have not been achieved. This research aims to identify the main challenges and obstacles in implementing DevOps at IPHO. A comprehensive examination of existing literature is employed to recognize prevalent difficulties encountered by organizations when implementing DevOps. The main factors are ranked using the Fuzzy Analytic Hierarchy Process (FAHP) based on survey data from DevOps practitioners at IPHO. This study helps fill in some gaps left by empirical studies on the challenges in applying DevOps, especially in the public healthcare sector. It also streamlines the data collection and analysis process by utilizing FAHP, simplifying the survey process, and reducing the number of questions compared to previous approaches. According to the research findings, the primary hurdle that requires attention is the mindset to transform from a traditional approach to continuous delivery. In addition, the lack of understanding about the benefits of implementing DevOps and the lack of cross-functional leadership are also identified as challenges that need to be considered. However, IPHO does not view the use of legacy tools and technologies as a significant impediment to adopting DevOps.

**Keywords**—DevOps; challenges; fuzzy AHP; software development

## I. INTRODUCTION

Organizations are currently competing to speed up the conversion of business requirements and business concepts into software applications. The rapidity of software application development plays a critical role in addressing the swiftly shifting business needs of customers and accommodating the ever-changing demands of the business landscape [1]. Software organizations must release effective and sustainable products in a volatile market to compete and maintain a competitive advantage [2]. Therefore, companies that focus on software development must continuously improve their project management practices to achieve higher product quality and enter the market faster [3].

The new approach known as DevOps is often described as a way to deliver software faster and with higher quality through collaboration between development (Dev) and operations (Ops) teams [3]. DevOps is still considered a relatively new approach in software engineering but has garnered significant attention from organizations seeking to improve their software delivery processes [1]. DevOps encompasses various aspects such as tools, organizational culture, practices, and collaboration and can help the software

industry achieve better performance and development processes [2].

As Indonesia's largest social insurance company, appointed by the Indonesian government to execute a social health insurance program [4], the Indonesian Public Health Organization (IPHO) is currently attempting to implement DevOps technology in its information system development process. However, the expected benefits have not been realized after running this program for three years. Although IPHO obtained DevOps software licenses and technical support in 2020 [5], no software has yet to utilize DevOps technology in intensive production successfully.

This research aims to identify the key factors that pose challenges and barriers to implementing DevOps at IPHO. The study begins with a literature review to identify challenges and obstacles commonly encountered by companies implementing DevOps technology and culture. A survey is then conducted among the teams at IPHO who have been directly involved in the DevOps implementation process over the past three years. From the survey, the factors posing the main challenges and barriers at IPHO are ranked using the Fuzzy Analytical Hierarchy Process (FAHP) approach.

Comprehensive empirical research on the analysis of challenges in implementing DevOps is currently scarce, especially in sectors such as Public Health Organizations. Hence, this research has the potential to contribute novelty to the domain of DevOps adoption. Furthermore, it will streamline the data collection and analysis of survey data by employing FAHP (Fuzzy Analytic Hierarchy Process). Previous research conducted by M. A. Akbar et al. [6] and A. A. Khan et al. [7] involved two stages of the survey: sentiment assessment of the categories for ranking and pairwise comparison survey. These previous approaches took a long time and required many questions of survey. In this study, the researcher attempts to conduct only one survey stage: a sentiment survey regarding the suitability of categories to real-life events. The researcher will then use the technique of geometric means to translate the scores provided by the survey respondents into Triangular Fuzzy Numbers (TFN) for the computation of pairwise comparisons. The study poses the following research issues regarding this:

RQ1: What are the common challenges and barriers in a company during the implementation of DevOps?

RQ2: What are the main challenges and barriers in implementing DevOps at IPHO?

In this research, the author proposes several sections to provide a clear and comprehensive understanding. Section II provides an in-depth discussion of the research's theoretical foundation, while Section III explains the research methodology used to collect and analyze data. Section IV presents the study results and discussion. Finally, Section V summarizes the study and provides limitations on this research and recommendations for future research directions.

## II. BACKGROUND OF DEVOPS

Evolution tools and methodologies have undergone considerable modifications as a result of the quick development of information technology. Businesses are driven to switch from manual to digital processes because automation can increase efficiency and ensure consistent product quality. The demand for software systems has dramatically expanded [8].

Software organizations continue to seek active development approaches to meet market demands by developing and delivering high-quality software on time and within budget [8]. Dörnenburg [9] suggests that software organizations must adopt new and efficient software development approaches to respond to market demands and effectively address technological changes.

Agile paradigms like Scrum and Kanban have superseded traditional software development methodologies like Waterfall and Spiral to keep up with technological changes and market trends. Because manual processing is prone to mistakes and can cause delays in feedback, production and operational processes have grown more complex [10]. Therefore, a new and more effective software development model known as DevOps has arisen to stay up with the current trend in the software business.

DevOps offers supplemental agile methods based on agile concepts and operational considerations. This strategy facilitates the rapid and continual delivery of developed features over shorter life cycles. DevOps initially had conflicting meanings in the software industry since some communities saw it as a career path requiring expertise in development and operations [11]. Research that defines DevOps as a development environment where growth and operational teams collaborate closely has resolved this issue [11]–[13].

Although there is still a division between developers and operations in DevOps, the operations team oversees changes made to service levels and production [10]. In contrast, the development team consistently creates new features to meet established business objectives. The two teams' tools, procedures, and knowledge bases are distinct. With the help of this system, the development team can continually add new features. In contrast, the operational team works to run the most recent version and control modifications to uphold project quality standards and other non-functional needs [10].

An automated pipeline is needed to address the information flow between the development and operations teams [14]. According to Humble and Farley's [10] automated deployment pipeline, each software version committed to the repository must be prepared for production. To provide a route that

enables automated development, testing, and the quick delivery of tested software features to production, Sten [15] highlights the significance of automation procedures. Callanan and Spillane [16] refer to the deployment pipeline as the DevOps platform and emphasize the need for continuous delivery.

## III. RESEARCH METHODOLOGY

The process depicted in Fig. 1 encompasses several stages. To identify the optimal strategies pertaining to DevOps initiatives, a systematic literature review (SLR) was conducted initially. Subsequently, a survey with a questionnaire was carried out to gather feedback and viewpoints from business experts regarding the selected DevOps concepts. The identified challenges were then prioritized using the fuzzy Analytic Hierarchy Process (fuzzy-AHP) method.

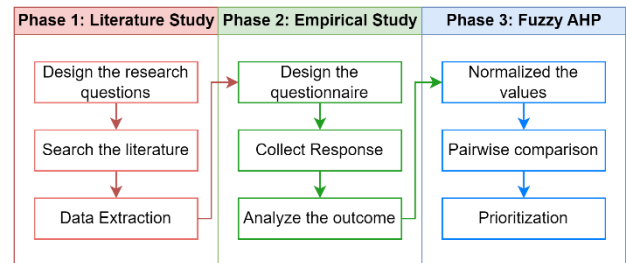


Fig. 1. The utilized research methodology.

- Phase 1: It entailed performing a comprehensive review of the literature to determine the difficulties with DevOps as stated in previous studies.
- Phase 2: It comprised a questionnaire survey aimed at validating the identified challenges empirically, specifically from an industrial perspective.
- Phase 3: Applied the fuzzy AHP methodology to determine the relative importance of the identified challenges in DevOps.

### A. Conducting a Systematic Literature Review (SLR)

The systematic literature review (SLR) used in this study was conducted following the standards set out by Kitchenham and Charter [16]. The SLR process was divided into three stages: planning, conducting, and reporting.

1) *Planning the review*: Making the protocols for data collection and analysis is part of planning. The procedures in the review methodology listed below are employed to obtain and evaluate the available literature to respond to the research question.

a) *Data Collection Source*: Finding literature pertinent to the study's purpose requires carefully selecting data sources. We followed the recommendations of Zhang [17] and Chen [18] in this investigation. The following digital archives found the primary studies relevant to the search:

- Science Direct at <https://www.sciencedirect.com>
- ACM Digital Library at <https://dl.acm.org>
- ProQuest at <https://www.proquest.com>

- Scopus at <https://www.scopus.com>
- IEEE Xplore at <https://ieeexplore.ieee.org>

b) *Search String*: We used the recommendations in the pertinent literature to create the search string for this investigation. First, essential phrases from the relevant papers were determined. Then, the search string was created by combining the "AND" and "OR" operators with the following key phrases and their synonyms: ("DevOps" OR "Development and Operation") AND ("challenge" OR "barriers" OR "obstacles" OR "hurdles" OR "difficulties" OR "impediments" OR "hindrance")

c) *Inclusion And Exclusion Criteria*: The protocols' primary function is to apply the exclusion and inclusion standards to literature found using search terms. Other information technology research, such as those by Niazi and colleagues [19] and the work of Akbar [20], have also used this strategy. The requirements for inclusion are specified in the procedures below:

- Inclusion Criteria:
  - The publication must be from a credible journal, conference, or book.
  - The publication ought to go into the difficulties of putting DevOps into practice.
  - The report must clearly explain how DevOps is implemented.
  - English must be used in the writing of the chosen article.
- Exclusion Criteria:
  - Only the most comprehensive one will be considered when two studies are relevant to the same project.
  - The paper lacks specific information regarding the implementation of DevOps.
  - Studies unrelated to DevOps will be disregarded.
  - Literature studies will not be considered.

d) *Conducting Quality Assessment (QA)*: The efficacy of the chosen literature in answering the study aim was evaluated using the quality assessment (QA) procedure. The QA procedure adhered to the recommendations made in [16]. The Likert scale is depicted in Table I and was used to evaluate the five questions created. Appendix B (Table XIV) has the full QA scores.

TABLE I. CHECKLIST FOR QA OF THE CHOSEN STUDIES

No	Checklist Questions	Likert scale
QA1	Does the analysis strategy use to answer the questions posed?	"Yes=1, Partial=0.5, NO=0"
QA2	Does the analysis look at the difficulties that come with DevOps?	"Yes=1, Partial=0.5, NO=0"
QA3	Does the report offer a convincing justification for putting DevOps into practice?	"Yes=1, Partial=0.5, NO=0"
QA4	Are the data gathered pertinent for using DevOps techniques?	"Yes=1, Partial=0.5, NO=0"
QA5	Do the outcomes that have been discovered support the research issues?	"Yes=1, Partial=0.5, NO=0"

2) *Conducting the Review*: In the initial response of the search string on the chosen databases, 906 studies were recovered. The gathered literature was further improved using the tollgate technique created by Afzal [21]. The tollgate technique involves five steps, and each meticulously carried out with the ultimate goal of selecting the studies for data extraction. As indicated in Fig. 2, 24 studies were chosen for the last data extraction procedure (see Table XIII in Appendix A).

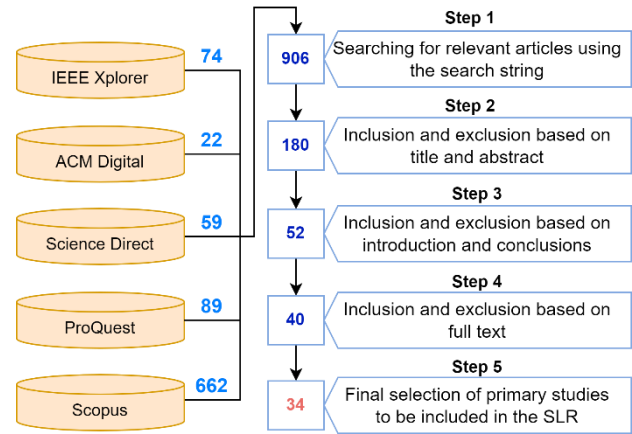


Fig. 2. Filtering formal literature.

a) *Assessment of Selected Study Quality*: The quality evaluation aims to analyze how well the chosen literature addresses the study's research topic. The selected studies are sufficiently relevant in addressing the research topic since 65% of the papers evaluated had a score of or higher than 60% (see Table XIV in Appendix B). During the quality evaluation procedure, a cutoff point of 50% was selected.

b) *Years of Publication of Chosen Articles*: During the data extraction stage, we gathered the studies' publication years to investigate the prevalence of DevOps literature. According to the frequency evaluation, the chosen studies cover 2018 through 2023, demonstrating a developing trend in DevOps research. The result indicates that the field of software engineering research is actively interested in DevOps.

### B. Conducting an Empirical Study

The questionnaire was created using the Google Forms platform, specifically the docs.google.com/forms platform. The questionnaire was broken up into three different sections for clarity and organization.

The first section's primary goal is to collect bibliographic data from survey respondents to give their comments some context. A set of closed-ended questions aimed at addressing the DevOps difficulties discovered via a thorough and systematic literature review (SLR) research was added in the second part. These closed-ended questions provide respondents with a predetermined range of response alternatives, enabling a complete examination of the problems that have been highlighted.



Finally, the questionnaire's third section includes open-ended questions meant to elicit participants' responses on any additional DevOps security problems that the SLR research might not have covered. Therefore, contributors are encouraged to offer thorough justifications, viewpoints, and proposed solutions in this part to contribute to a thorough knowledge of the topic.

The survey's target population includes all IPHO personnel actively working on the DevOps implementation project, which began in 2020 at the early stages of DevOps adoption and will continue until this research is finished in 2023. The main goal is to have a more profound knowledge of the limitations and difficulties encountered while IPHO embraced DevOps.

### C. Utilizing The Fuzzy Analytic Hierarchy Process (Fuzzy AHP)

The fuzzy AHP presents a practical approach for addressing multicriteria decision-making problems. One of the key advantages of utilizing fuzzy AHP is its ease of application and comprehensibility, making it accessible to users. Moreover, it can handle both quantitative and qualitative data effectively. The primary steps employed in implementing the fuzzy AHP methodology are as follows:

- Step 1: Organize the complex decision-making problem into a hierarchy.
- Step 2: Determine the highest and lowest values for each hierarchy element.
- Step 3: Check each pairwise comparison matrix's consistency to confirm correctness.

- Step 4: Establish final ranks for each factor and its corresponding categories.

When assessing the relative importance of various criteria, the traditional Analytic Hierarchy Process (AHP) cannot deal with the ambiguity and vagueness of decision-makers. The AHP approach has been combined with fuzzy theory to solve this issue, creating a fuzzy AHP. As mentioned in the reference [22], this method permits the determination of more precise and dependable judgments in real-time and unforeseen issue scenarios.

In multicriteria decision-making (MCDM) situations, fuzzy AHP is applicable to both qualitative as well as quantitative inputs. The extent analysis approach is used in this method to estimate the priority weight of certain criteria and express preference ratings for the criteria utilizing triangular fuzzy numbers. For this investigation, we used Chang's fuzzy AHP method, which produces more accurate and reliable findings than the traditional AHP method [23].

## IV. RESULT AND DISCUSSION

### A. Findings of SLR Study

The SLR technique was carefully used to pinpoint DevOps operations' challenging and essential elements. Table II lists the 34 tasks that were examined in total. The five categories of "Culture," "Management," "Process," "Teams," and "Tools," which are seen to be crucial elements for the effective use of DevOps principles in the software industry, were then used to map these challenges. This procedure led to the creation of the structure shown in Fig. 3. This mapping exercise's primary goal is to run a fuzzy AHP analysis.

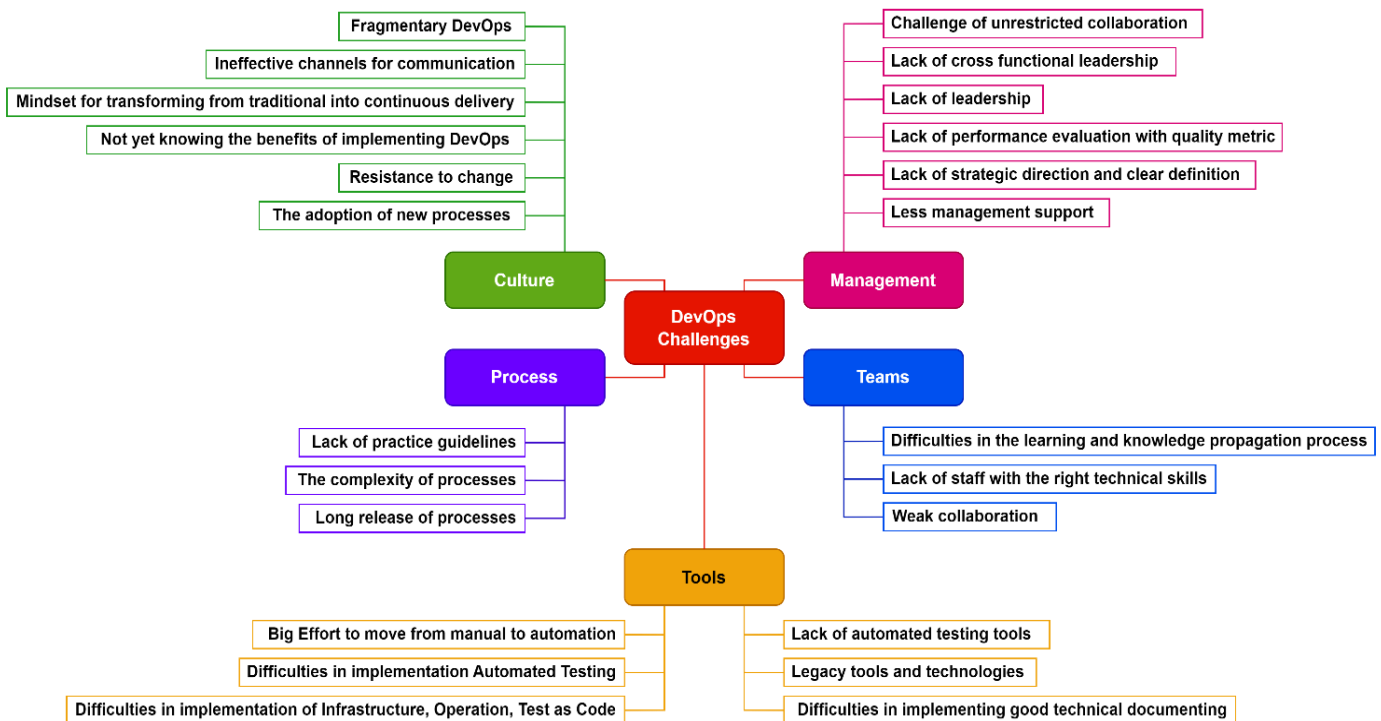


Fig. 3. Mapping of investigated challenges into categories.

TABLE II. LIST OF DEVOPS CHALLENGES

No	Factors	Source
C1	Fragmentary DevOps	[24]–[26]
C2	Ineffective channels for communication	[27]
C3	Mindset for transforming from traditional into continuous delivery	[28]–[30]
C4	Not yet knowing the benefits of implementing DevOps	[28], [31]
C5	Resistance to change	[3], [32]
C6	The adoption of new processes	[3]
C7	Difficulties in allocating resources	[33]–[36]
C8	Lack of cross-functional leadership	[25]
C9	Lack of leadership	[25], [37]
C10	Lack of performance evaluation with a quality metric	[31], [37]–[40]
C11	Lack of a clear concept and strategic direction	[24], [37], [41]
C12	Less management support	[25], [27]
C13	Lack of practice guidelines	[24]
C14	The complexity of processes	[26], [29]
C15	Long release of processes	[29], [42]
C16	Difficulties in the learning and knowledge propagation process	[32]
C17	Lack of technical expertise staff	[25], [27], [29], [41]
C18	Weak collaboration	[24], [25], [27], [41]
C19	Big effort to move from manual to automation	[29], [33], [43]
C20	Difficulties in the implementation of Automated Testing	[34]
C21	Difficulties in implementation of Infrastructure, Operation, Test as Code	[29], [34], [35], [44]
C22	Difficulties in implementing good technical documenting	[35]
C23	Lack of automated testing tools	[34], [45]
C24	Legacy tools and technologies	[26], [29], [37]

**B. Empirical Study Results**

1) *Analysis of survey participants' demographic data:* Detailed demographic data on the survey respondents was gathered while the data was being collected. According to Patten [46], demographic information is critical for understanding survey respondents and assessing if the participants in a given research are a representative sample of the target population to generalize the findings. According to Finstad [47], bibliographic information on survey respondents might provide insight into the maturity of the gathered dataset. Furthermore, Altman [48] underlined that knowing more about survey respondents aids in comprehending the target population's viewpoints. This study acknowledged the significance of the respondents' bibliographic information, and an analysis of many factors, including respondents'

designation and organization size, was done. The following sections explain the findings of this analysis.

a) *Designation of the respondent:* The importance of the influencing elements, which differ depending on the characteristics of the respondents, were stressed by Finstad [47]. Niazi [19] adds that the practitioner's position affects how a factor has an effect, adding that the influence of a component may be appropriately rated if the responder regularly works with that specific issue. The analysis of respondents based on their titles is shown in Fig. 4, which demonstrates that project managers make up most survey respondents. According to the results, "Junior Programmer (Staff)", "Senior Programmer (Assistant Manager)", and "Senior Programmer (Manager)" are the most often used response designations.

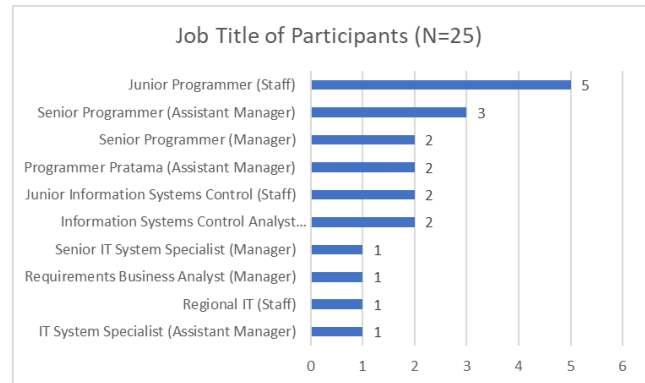


Fig. 4. Job title evaluation in survey.

b) *Respondent's Experience:* An analysis was conducted on the experience of the survey participants. The median and mean values were computed, resulting in scores of 3 and 2.4, respectively, indicating a relatively young group of participants. Additionally, notable variations in the respondents' experience were briefly observed. Fig. 5 shows a visual depiction of the survey respondents' information.

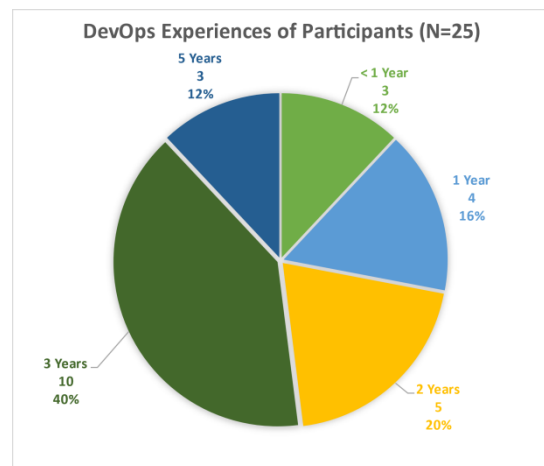


Fig. 5. Survey respondents' experiences.

TABLE III. AN EMPIRICAL INVESTIGATION OF CHALLENGES FACTORS

ID	Challenges List	Number of Responses (P=25)							
		Positive			Negative			Neutral	
		S-A	A	%	D	S-D	%	N	%
<b>P1</b>	<b>Culture</b>	<b>5</b>	<b>15</b>	<b>74%</b>	<b>0</b>	<b>0</b>	<b>0%</b>	<b>7</b>	<b>26%</b>
C1	Fragmentary DevOps	12	7	70%	7	0	28%	1	2%
C2	Ineffective channels for communication	8	14	81%	0	0	0%	5	19%
C3	Mindset for transforming from traditional into continuous delivery	13	11	89%	1	0	4%	2	7%
C4	Not yet knowing the benefits of implementing DevOps	16	9	94%	2	0	7%	0	0%
C5	Resistance to change	10	5	56%	6	1	26%	5	19%
C6	The adoption of new processes	6	12	67%	0	3	9%	6	24%
<b>P2</b>	<b>Management</b>	<b>2</b>	<b>20</b>	<b>81%</b>	<b>0</b>	<b>0</b>	<b>0%</b>	<b>5</b>	<b>19%</b>
C7	Difficulties for allocating resources	7	10	63%	3	1	15%	5	19%
C8	Lack of cross functional leadership	12	13	93%	0	0	0%	2	7%
C9	Lack of leadership	9	16	93%	1	0	4%	1	4%
C10	Lack of performance evaluation with quality metric	6	11	63%	2	2	15%	6	22%
C11	Lack of strategic direction and clear definition	9	11	74%	2	1	11%	4	15%
C12	Less management support	8	8	59%	3	1	15%	7	26%
<b>P3</b>	<b>Process</b>	<b>5</b>	<b>12</b>	<b>63%</b>	<b>1</b>	<b>0</b>	<b>4%</b>	<b>9</b>	<b>33%</b>
C13	Lack of practice guidelines	9	15	89%	1	0	4%	2	7%
C14	The complexity of processes	7	11	67%	4	2	22%	3	11%
C15	Long release of processes	4	14	67%	5	1	22%	3	11%
<b>P4</b>	<b>Teams</b>	<b>5</b>	<b>12</b>	<b>63%</b>	<b>0</b>	<b>0</b>	<b>0%</b>	<b>10</b>	<b>37%</b>
C16	Difficulties in the learning and knowledge propagation process	10	11	78%	1	0	4%	5	19%
C17	Lack of staff with the right technical skills	7	12	70%	2	0	7%	6	22%
C18	Weak collaboration	6	11	63%	5	0	19%	5	19%
<b>P5</b>	<b>Tools</b>	<b>4</b>	<b>13</b>	<b>63%</b>	<b>0</b>	<b>0</b>	<b>0%</b>	<b>10</b>	<b>37%</b>
C19	Big effort to move from manual to automation	11	8	70%	4	0	15%	4	15%
C20	Difficulties in implementation Automated Testing	11	10	78%	2	0	7%	4	15%
C21	Difficulties in implementation of Infrastructure, Operation, Test as Code	10	8	67%	4	0	15%	5	19%
C22	Difficulties in implementing good technical documenting	9	13	81%	1	0	4%	4	15%
C23	Lack of automated testing tools	9	9	67%	3	0	11%	6	22%
C24	Legacy tools and technologies	6	11	63%	5	2	27%	3	10%
<b>Average</b>				<b>73%</b>			<b>10%</b>		<b>17%</b>

2) *Analysis of responses to DevOps challenges:* The empirical study's primary objective was to learn more from business experts about the difficulties faced by DevOps, as determined by a systematic literature review (SLR). Three categories—positive (“agree, strongly agree”), negative (“disagree, strongly disagree”), and “neutral”—were used to classify the replies given by practitioners. The positive category represents the proportion of survey participants who acknowledged the difficulties that might negatively influence DevOps techniques. The opposing group comprises respondents who frequently disagreed with the challenges noted in the SLR research. The neutral type represents the participants who often expressed uncertainty about how the specified factors will affect DevOps operations. Please see Table III for more specific data.

The study's findings are shown in Table III, which shows that most survey respondents concur that DevOps has a bad relationship with the issues in actual operations. According to the frequency analysis, over 50% of survey respondents considered each challenging element. C4, or “Not yet knowing the benefits of implementing DevOps,” was cited as the most challenging problem by survey respondents (94%).

The poll respondents named P2 (Management, 81%) as the most critical categories among the researched complex variables, with P1 (Culture, 74%) placing second. The third most important types of problems are P3 (Process, 63%) and P4 (Team, 73%).

Among the challenges categorized as negative factors, C1(Fragmentary DevOps) emerged as the highest-ranking challenge, with 28% of the respondents disagreeing with its classification as a significant factor in DevOps practices. Following closely behind, and C24 (Legacy tools and

technologies, 27%) received the second highest level of disagreement among the respondents.

Additionally, challenges were sorted based on respondents' limited understanding of their impact on both DevOps implementation with neutral response options. The top-ranked challenge, labeled as C6 (The adoption of new processes), received a 24% response rate. It was closely followed by C10 (Lack of performance evaluation with quality metric), C17 (Lack of staff with the right technical skills), and C23 (Lack of automated testing tools), all of which received a 22% response rate and were ranked as the second-highest challenges.

C. Implementing Fuzzy AHP

The fuzzy-AHP used to investigate the rank of challenges and the categories they fall into is presented in this part. The issues were prioritized using the fuzzy AHP step-by-step procedures mentioned in the preceding section.

Step 1 (Hierarchical Categorization of Complex Problems): The complicated problem is separated into linked decision-making components using the method described in [49] and [50] to do the fuzzy AHP analysis. The top level of the problem reflects the primary goal, whereas Stages 2 and 3 show the types of issues and their accompanying challenges. Fig. 6 depicts the suggested hierarchical structure.

Step 2 (The process of pairwise comparisons): On the basis of professional judgments, the pairwise comparison was undertaken. The author continued using the first questionnaire's data (see Table XV in Appendix C) to gather reference values for the paired comparison data process utilizing the geometric mean calculation approach to provide sufficient and quicker pairwise comparison data. Triangular fuzzy numbers (TFNs) can be generated using the geometric mean from survey respondents' judgments. The following geometric mean formula was applied in this study:

$$\text{Geometric mean} = \sqrt[n]{j_1 \times j_2 \times j_3 \dots j_n} \tag{1}$$

$j$  = Individual judgment weights  
 $n$  = Count of judgments

Step 3 (Verifying the Pairwise Matrix's Consistency): The method for determining if the pairwise comparison matrices are consistent is laid out sequentially in this section. The matrices must show consistency in fuzzy AHP. Consideration is given to the categories on the Likert scale (Table IV). To create the matching Fuzzy Crisp Matrix (FCM) displayed in Table V, the primary categories triangular fuzzy numbers in the pairwise comparison matrix are defuzzification into crisp numbers.

The fuzzy Analytic Hierarchy Process (AHP) can use a comparison matrix that is adequate and trustworthy for many issue categories. The consistency ratio for each category of challenge elements was computed using the same approach, and the results are shown according to Tables VI to X in the appropriate order.

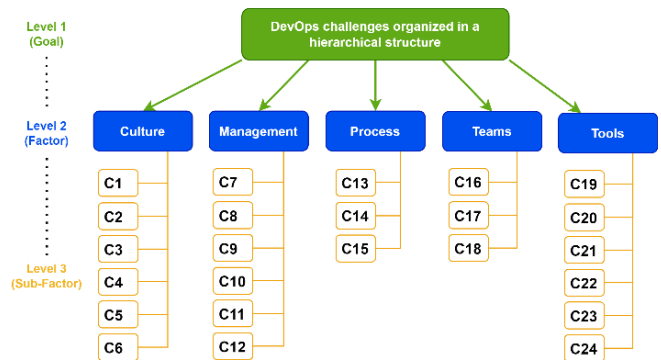


Fig. 6. The hierarchy of the problem structure.

TABLE IV. CONVERSION SCALE OF A FUZZY [51]

Grade	Linguistic scale	Triangular fuzzy scale	Triangular fuzzy reciprocal scale
1	Element j holds the same level of importance as element i.	(1, 1, 1)	(1, 1, 1)
2	Slightly of lower importance.	(1, 2, 3)	(0.33, 0.50, 1.00)
3	Somewhat important, falling within the range of slightly important.	(2, 3, 4)	(0.25, 0.33, 0.50)
4	Moderately important, ranging from slightly important to more important.	(3, 4, 5)	(0.20, 0.25, 0.33)
5	Significantly more important.	(4, 5, 6)	(0.17, 0.20, 0.25)
6	Between a more important and highly important.	(5, 6, 7)	(0.14, 0.17, 0.20)
7	Highly important.	(6, 7, 8)	(0.13, 0.14, 0.17)
8	Between highly important and most important	(7, 8, 9)	(0.11, 0.13, 0.14)
9	Most significance.	(9, 9, 9)	(0.11, 0.11, 0.11)

TABLE V. CHALLENGE CATEGORY FUZZY-CRISP MATRIX (FCM)

	P1	P2	P3	P4	P5	Priority Vector Weight
P1	1.00	3.00	8.00	5.00	6.00	0.4995
P2	0.33	1.00	6.00	3.00	4.00	0.2585
P3	0.13	0.17	1.00	0.25	0.33	0.0388
P4	0.20	0.33	4.00	1.00	2.00	0.1216
P5	0.17	0.25	3.00	0.50	1.00	0.0816

TABLE VI. CULTURE CATEGORY PAIRWISE COMPARISONS

	C1	C2	C3	C4	C5	C6
C1	(1.00, 1.00, 1.00)	(0.33, 0.50, 1.00)	(0.17, 0.20, 0.25)	(0.20, 0.25, 0.33)	(3.00, 4.00, 5.00)	(2.00, 3.00, 4.00)
C2	(1.00, 2.00, 3.00)	(1.00, 1.00, 1.00)	(0.20, 0.25, 0.33)	(0.25, 0.33, 0.50)	(4.00, 5.00, 6.00)	(3.00, 4.00, 5.00)
C3	(4.00, 5.00, 6.00)	(3.00, 4.00, 5.00)	(1.00, 1.00, 1.00)	(1.00, 2.00, 3.00)	(7.00, 8.00, 9.00)	(6.00, 7.00, 8.00)
C4	(3.00, 4.00, 5.00)	(2.00, 3.00, 4.00)	(0.33, 0.50, 1.00)	(1.00, 1.00, 1.00)	(6.00, 7.00, 8.00)	(5.00, 6.00, 7.00)
C5	(0.20, 0.25, 0.33)	(0.17, 0.20, 0.25)	(0.11, 0.13, 0.14)	(0.13, 0.14, 0.17)	(1.00, 1.00, 1.00)	(0.33, 0.50, 1.00)
C6	(0.25, 0.33, 0.50)	(0.20, 0.25, 0.33)	(0.13, 0.14, 0.17)	(0.14, 0.17, 0.20)	(1.00, 2.00, 3.00)	(1.00, 1.00, 1.00)

I<sub>max</sub> = 6.222, CI= 0.044, CR= 0.036

TABLE VII. MANAGEMENT CATEGORY PAIRWISE COMPARISONS

	C7	C8	C9	C10	C11	C12
C7	(1.00, 1.00, 1.00)	(0.20, 0.25, 0.33)	(0.25, 0.33, 0.50)	(4.00, 5.00, 6.00)	(1.00, 2.00, 3.00)	(3.00, 4.00, 5.00)
C8	(3.00, 4.00, 5.00)	(1.00, 1.00, 1.00)	(1.00, 2.00, 3.00)	(7.00, 8.00, 9.00)	(4.00, 5.00, 6.00)	(6.00, 7.00, 8.00)
C9	(2.00, 3.00, 4.00)	(0.33, 0.50, 1.00)	(1.00, 1.00, 1.00)	(6.00, 7.00, 8.00)	(3.00, 4.00, 5.00)	(5.00, 6.00, 7.00)
C10	(0.17, 0.20, 0.25)	(0.11, 0.13, 0.14)	(0.13, 0.14, 0.17)	(1.00, 1.00, 1.00)	(0.20, 0.25, 0.33)	(0.33, 0.50, 1.00)
C11	(0.33, 0.50, 1.00)	(0.17, 0.20, 0.25)	(0.20, 0.25, 0.33)	(3.00, 4.00, 5.00)	(1.00, 1.00, 1.00)	(2.00, 3.00, 4.00)
C12	(0.20, 0.25, 0.33)	(0.13, 0.14, 0.17)	(0.14, 0.17, 0.20)	(1.00, 2.00, 3.00)	(0.25, 0.33, 0.50)	(1.00, 1.00, 1.00)

I<sub>max</sub> = 6.112, CI= 0.039, CR= 0.031

TABLE VIII. PROCESS CATEGORY PAIRWISE COMPARISONS

	C13	C14	C15
C13	(1.00, 1.00, 1.00)	(6.00, 7.00, 8.00)	(3.00, 4.00, 5.00)
C14	(0.13, 0.14, 0.17)	(1.00, 1.00, 1.00)	(0.20, 0.25, 0.33)
C15	(0.20, 0.25, 0.33)	(3.00, 4.00, 5.00)	(1.00, 1.00, 1.00)

I<sub>max</sub> = 3.076, CI= 0.038, CR= 0.066

TABLE IX. TEAMS CATEGORY PAIRWISE COMPARISONS

	C16	C17	C18
C16	(1.00, 1.00, 1.00)	(3.00, 4.00, 5.00)	(6.00, 7.00, 8.00)
C17	(0.20, 0.25, 0.33)	(1.00, 1.00, 1.00)	(3.00, 4.00, 5.00)
C18	(0.13, 0.14, 0.17)	(0.20, 0.25, 0.33)	(1.00, 1.00, 1.00)

I<sub>max</sub> = 2.023, CI= 0.025, CR= 0.043

TABLE X. TOOLS CATEGORY PAIRWISE COMPARISONS

	C19	C20	C21	C22	C23	C24
C19	(1.00, 1.00, 1.00)	(0.14, 0.17, 0.20)	(0.20, 0.25, 0.33)	(0.13, 0.14, 0.17)	(0.25, 0.33, 0.50)	(1.00, 2.00, 3.00)
C20	(5.00, 6.00, 7.00)	(1.00, 1.00, 1.00)	(2.00, 3.00, 4.00)	(0.33, 0.50, 1.00)	(3.00, 4.00, 5.00)	(6.00, 7.00, 8.00)
C21	(3.00, 4.00, 5.00)	(0.25, 0.33, 0.50)	(1.00, 1.00, 1.00)	(0.20, 0.25, 0.33)	(1.00, 2.00, 3.00)	(4.00, 5.00, 6.00)
C22	(6.00, 7.00, 8.00)	(1.00, 2.00, 3.00)	(3.00, 4.00, 5.00)	(1.00, 1.00, 1.00)	(4.00, 5.00, 6.00)	(7.00, 8.00, 9.00)
C23	(2.00, 3.00, 4.00)	(0.20, 0.25, 0.33)	(0.33, 0.50, 1.00)	(0.17, 0.20, 0.25)	(1.00, 1.00, 1.00)	(3.00, 4.00, 5.00)
C24	(0.33, 0.50, 1.00)	(0.13, 0.14, 0.17)	(0.17, 0.20, 0.25)	(0.11, 0.13, 0.14)	(0.20, 0.25, 0.33)	(1.00, 1.00, 1.00)

I<sub>max</sub> = 6.217, CI= 0.022, CR= 0.018

Tables VI to X indicate that the consistency ratio (CR) is below 0.1, allowing the questionnaire data to be used for the subsequent step, calculating local and global ranking weights.

Step 4 (Weights are determined locally and globally): The weights of the challenges and their associated categories, both locally and globally, were computed. Table XI displays the findings and compares each task's importance to all other challenges (global weight), showing how each problem ranks within its category.

TABLE XI. CALCULATE THE CUMULATIVE WEIGHT OF THE CHALLENGES

Category Weight	Challenges	Local		Global	
		Weight	Rank	Weight	Rank
Culture (0.49952)	C1	0.09543	4	0.04767	7
	C2	0.13911	3	0.06949	6
	C3	0.39976	1	0.19968	1
	C4	0.28827	2	0.14400	2
	C5	0.03210	6	0.01603	15
	C6	0.04534	5	0.02265	14

Category Weight	Challenges	Local		Global	
		Weight	Rank	Weight	Rank
Management (0.25854)	C7	0.13911	3	0.03597	8
	C8	0.39976	1	0.10335	3
	C9	0.28827	2	0.07453	5
	C10	0.03210	6	0.00830	20
	C11	0.09543	4	0.02467	12
	C12	0.04534	5	0.01172	16
Process (0.03880)	C13	0.69183	1	0.02685	11
	C14	0.07648	3	0.00297	23
	C15	0.23169	2	0.00899	19
Teams (0.08156)	C16	0.69183	1	0.08411	4
	C17	0.23169	2	0.02817	10
	C18	0.07648	3	0.00930	18
Tools (0.12158)	C19	0.04534	5	0.00370	22
	C20	0.28827	2	0.02351	13
	C21	0.13911	3	0.01134	17
	C22	0.39976	1	0.03260	9
	C23	0.09543	4	0.00778	21
	C24	0.03210	6	0.00262	24

Step 5 (Challenges Prioritization): The fuzzy AHP analysis's primary goal is to rank the researched challenges according to their importance to the DevOps paradigm. Table XII lists the final standings for each challenge.

TABLE XII. PRIORITY ORDER FOR THE DIFFICULTIES

ID	Challenges List	Rank
C3	Mindset for transforming from traditional into continuous delivery	1
C4	Not yet knowing the benefits of implementing DevOps	2
C8	Lack of cross-functional leadership	3
C16	Difficulties in the learning and knowledge propagation process	4
C9	Lack of leadership	5
C2	Ineffective channels for communication	6
C1	Fragmentary DevOps	7
C7	Difficulties in allocating resources	8
C22	Difficulties in implementing good technical documenting	9
C17	Lack of staff with the right technical skills	10
C13	Lack of practice guidelines	11
C11	Lack of strategic direction and clear definition	12
C20	Difficulties in the implementation of Automated Testing	13
C6	The adoption of new processes	14
C5	Resistance to change	15
C12	Less management support	16
C21	Difficulties in implementation of Infrastructure, Operation, Test as Code	17
C18	Weak collaboration	18
C15	Long release of processes	19
C10	Lack of performance evaluation with quality metric	20
C23	Lack of automated testing tools	21
C19	Big effort to move from manual to automation	22
C14	The complexity of processes	23
C24	Legacy tools and technologies	24

Based on the global weights, it is decided that C3 (Mindset for changing from conventional into continuous delivery) is the most critical challenge that must be solved to implement DevOps methods at IPHO effectively. Additionally, C4 (Not understanding the advantages of applying DevOps) and C8 (Lack of cross-functional leadership) are listed as the second and third most significant priority difficulties for implementing DevOps methods. It is also important to note that C24 (Legacy tools and technologies) is listed as the least major issue for the DevOps paradigm in IPHO.

## V. CONCLUSION

IPHO has adopted DevOps principles due to the significance of expediting software application development to satisfy quickly changing business demands and preserve a competitive edge. The anticipated advantages have not materialized despite this program being in place for three years. This study tries to pinpoint the key variables that IPHO must overcome to use DevOps technology successfully.

The research utilizes a systematic literature review to identify common challenges companies face when implementing DevOps technology and culture. Furthermore, over the past three years, a survey has been conducted among teams directly involved in the DevOps implementation process at IPHO. Using the Fuzzy Analytical Hierarchy Process

(FAHP) method, the detected elements that provide significant problems and impediments are rated.

The limits of empirical studies concerning difficulties in DevOps adoption, particularly in the public healthcare sector, are addressed in this study. It also makes data gathering and analysis easier by employing FAHP, which streamlines the survey process and uses fewer questions than earlier methods.

The most critical challenge that has to be resolved for IPHO to apply DevOps methods effectively is C3 (Mindset for transitioning from conventional to continuous delivery), according to the global weights used in the analysis. In addition, C4 (Lack of knowledge about the advantages of applying DevOps) and C8 (Lack of cross-functional leadership) are noted as the second and third priority hurdles in implementing DevOps methods. It is important to note that at IPHO, C24 (Legacy tools and technologies) is rated as the least important issue for the DevOps concept.

## A. Limitations

Although this study has implemented a comprehensive research methodology, the authors acknowledge the presence of limitations. One of these limitations is the constraint of using only one public health institution as a case study, resulting in a study with a limited scope and more relevant to that specific institution. Additionally, the number of respondents involved in this research is considered inadequate for achieving a more robust analysis.

## B. Future Work

In the future, researchers can still conduct multivocal literature studies to examine the factors influencing DevOps practices but expand the research scope to investigate the success and challenges of adopting DevOps in organizations and involve multiple companies for more comprehensive results. Additionally, utilizing multiple case studies within the same sector can provide a larger sample size and generate more comprehensive analyses.

## ACKNOWLEDGMENT

This research was carried out with the support of a fully funded scholarship program by IPHO, which the first author obtained for study at the University of Indonesia.

## REFERENCES

- [1] M. Lazuardi, T. Raharjo, B. Hardian, and T. Simanungkalit, "Perceived Benefits of DevOps Implementation in Organization: A Systematic Literature Review," *ACM Int. Conf. Proceeding Ser.*, pp. 10–16, 2021, doi: 10.1145/3512716.3512718.
- [2] N. Azad and S. Hyrynsalmi, "DevOps critical success factors — A systematic literature review," *Inf. Softw. Technol.*, vol. 157, no. January, p. 107150, 2023, doi: 10.1016/j.infsof.2023.107150.
- [3] A. Trigo, J. Varajão, and L. Sousa, "DevOps adoption: Insights from a large European Telco," *Cogent Eng.*, vol. 9, no. 1, 2022, doi: 10.1080/23311916.2022.2083474.
- [4] Presiden Republik Indonesia, "Undang-Undang Republik Indonesia Nomor 24 Tahun 2011 Tentang Badan Penyelenggara Jaminan Sosial," 2011.
- [5] B. Kesehatan, "Laporan Kickoff Pengadaan DevOps BPJS Kesehatan," 2020.

- [6] M. A. Akbar et al., "Prioritization Based Taxonomy of DevOps Challenges Using Fuzzy AHP Analysis," *IEEE Access*, vol. 8, pp. 202487–202507, 2020, doi: 10.1109/ACCESS.2020.3035880.
- [7] A. A. Khan, M. Shameem, R. R. Kumar, S. Hussain, and X. Yan, "Fuzzy AHP based prioritization and taxonomy of software process improvement success factors in global software development," *Appl. Soft Comput. J.*, vol. 83, p. 105648, 2019, doi: 10.1016/j.asoc.2019.105648.
- [8] I. M. Sebastian, K. G. Moloney, J. W. Ross, N. O. Fonstad, C. Beath, and M. Mocker, "How big old companies navigate digital transformation," *MIS Q. Exec.*, vol. 16, no. 3, pp. 197–213, 2017, doi: 10.4324/9780429286797-6.
- [9] E. Dornenburg, "The Path to DevOps," *IEEE Softw.*, vol. 35, no. 5, pp. 71–75, 2018, doi: 10.1109/MS.2018.290110337.
- [10] J. Humble and D. Farley, *Continuous Delivery: Reliable Software Releases through Build, Test, and Deployment Automation*, 1st ed. Addison-Wesley Professional, 2010.
- [11] M. Senapathi, J. Buchan, and H. Osman, "DevOps capabilities, practices, and challenges: Insights from a case study," in *ACM International Conference Proceeding Series*, 2018, vol. Part F1377. doi: 10.1145/3210459.3210465.
- [12] J. Roche, "Adopting devops practices in quality assurance," *Commun. ACM*, vol. 56, no. 11, pp. 38–43, 2013, doi: 10.1145/2524713.2524721.
- [13] R. Jabbari, N. Bin Ali, K. Petersen, and B. Tanveer, "What is DevOps? A systematic mapping study on definitions and practices," *ACM Int. Conf. Proceeding Ser.*, vol. 24-May-201, no. March 2018, 2016, doi: 10.1145/2962695.2962707.
- [14] E. Woods, "Operational: The Forgotten Architectural View," *IEEE Softw.*, vol. 33, no. 3, pp. 20–23, 2016, doi: 10.1109/MS.2016.86.
- [15] Sten Pittet, "Continuous integration vs. delivery vs. deployment," *Atlassian.com*. <https://www.atlassian.com/continuous-delivery/principles/continuous-integration-vs-delivery-vs-deployment> (accessed Jun. 11, 2023).
- [16] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," 2007.
- [17] H. Zhang, M. A. Babar, and P. Tell, "Identifying relevant studies in software engineering," *Inf. Softw. Technol.*, vol. 53, no. 6, pp. 625–637, 2011, doi: 10.1016/j.infsof.2010.12.010.
- [18] L. Chen, M. A. Babar, and H. Zhang, "Towards an Evidence-Based Understanding of Electronic Data Sources," in *Proceedings of the 14th International Conference on Evaluation and Assessment in Software Engineering*, 2010, pp. 135–138.
- [19] M. Niazi, S. Mahmood, M. Alshayeb, A. M. Qureshi, K. Faisal, and N. Cerpa, "Toward successful project management in global software development," *Int. J. Proj. Manag.*, vol. 34, no. 8, pp. 1553–1567, 2016, doi: 10.1016/j.ijproman.2016.08.008.
- [20] M. A. Akbar et al., "Statistical Analysis of the Effects of Heavyweight and Lightweight Methodologies on the Six-Pointed Star Model," *IEEE Access*, vol. 6, pp. 8066–8079, 2018, doi: 10.1109/ACCESS.2018.2805702.
- [21] W. Afzal, R. Torkar, and R. Feldt, "A systematic review of search-based testing for non-functional system properties," *Inf. Softw. Technol.*, vol. 51, no. 6, pp. 957–976, 2009, doi: 10.1016/j.infsof.2008.12.005.
- [22] R. W. Saaty, "The analytic hierarchy process-what it is and how it is used," *Math. Model.*, vol. 9, no. 3–5, pp. 161–176, 1987, doi: 10.1016/0270-0255(87)90473-8.
- [23] D. Y. Chang, "Applications of the extent analysis method on fuzzy AHP," *Eur. J. Oper. Res.*, vol. 95, no. 3, pp. 649–655, 1996, doi: 10.1016/0377-2217(95)00300-2.
- [24] X. Zhou, H. Huang, H. Zhang, X. Huang, D. Shao, and C. Zhong, "A Cross-Company Ethnographic Study on Software Teams for DevOps and Microservices: Organization, Benefits, and Issues," *Proc. - Int. Conf. Softw. Eng.*, pp. 1–10, 2022, doi: 10.1109/ICSE-SEIP55303.2022.9794010.
- [25] K. Maroukian and S. R. Gulliver, "The Link between Transformational and Servant Leadership in DevOps-Oriented Organizations," in *ACM International Conference Proceeding Series*, 2020, pp. 21–29. doi: 10.1145/3393822.3432340.
- [26] J. Diaz, J. E. Perez, A. Yague, A. Villegas, and A. de Antona, "DevOps in Practice – A Preliminary Analysis of Two Multinational Companies," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11915 LNCS, pp. 323–330, 2019, doi: 10.1007/978-3-030-35333-9\_23.
- [27] A. Bijwe and P. Shankar, "Challenges of Adopting DevOps Culture on the Internet of Things Applications - A Solution Model," in *Proceedings of International Conference on Technological Advancements in Computational Sciences, ICTACS 2022*, 2022, pp. 638–645. doi: 10.1109/ICTACS56270.2022.9988182.
- [28] J. D'az, R. 'n Almaraz, J. P'erez, and J. Garbajosa, "DevOps in Practice: An Exploratory Case Study," *Proceedings of the 19th International Conference on Agile Software Development: Companion*, articleno = 1, numpages = 3. Association for Computing Machinery, 2018. doi: 10.1145/3234152.3234199.
- [29] R. K. Gupta, M. Venkatachalapathy, and F. K. Jeberla, "Challenges in Adopting Continuous Delivery and DevOps in a Globally Distributed Product Team: A Case Study of a Healthcare Organization," in *Proceedings - 2019 ACM/IEEE 14th International Conference on Global Software Engineering, ICGSE 2019*, 2019, pp. 30–34. doi: 10.1109/ICGSE.2019.00020.
- [30] A. B. Farid, Y. M. Helmy, and M. M. Bahloul, "Enhancing Lean Software Development by using Devops Practices," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 7, 2017, doi: 10.14569/IJACSA.2017.080736.
- [31] A. Lapointe-Boisvert, S. Mosser, and S. Trudel, "Towards Modelling Acceptance Tests as a Support for Software Measurement," in *Companion Proceedings - 24th International Conference on Model-Driven Engineering Languages and Systems, MODELS-C 2021*, 2021, pp. 827–832. doi: 10.1109/MODELS-C53483.2021.00129.
- [32] P. Mauro Lourenço, S. Mónica Ferreira da, and A. Leonardo Guerreiro, "DevOps Adoption: Eight Emergent Perspectives." *Cornell University Library, arXiv.org, Ithaca*, 2021.
- [33] H. Li, T.-H. P.-H. Chen, A. E. Hassan, M. Nasser, and P. Flora, "Adopting autonomic computing capabilities in existing large-scale systems: An industrial experience report," in *Proceedings - International Conference on Software Engineering*, 2018, pp. 1–10. doi: 10.1145/3183519.3183544.
- [34] M. Shahin, A. R. Nasab, and M. A. Babar, "A Qualitative Study of Architectural Design Issues in DevOps," *arXiv.org, Cornell University Library, arXiv.org PP - Ithaca, Ithaca, Nov. 13, 2021*.
- [35] J. Sorgalla, P. Wizenty, F. Rademacher, S. Sachweh, and A. Zündorf, "Applying Model-Driven Engineering to Stimulate the Adoption of DevOps Processes in Small and Medium-Sized Development Organizations," *arXiv.org, Cornell University Library, arXiv.org PP - Ithaca, Ithaca, Jul. 26, 2021*.
- [36] S. Bahaa, A. Z. Ghalwash, and H. Harb, "DataOps Lifecycle with a Case Study in Healthcare," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 1, 2023, doi: 10.14569/IJACSA.2023.0140115.
- [37] D. A. Meedeniya, I. D. Rubasinghe, and I. Perera, "Software artefacts consistency management towards continuous integration: A roadmap," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 4, pp. 100–110, 2019, doi: 10.14569/ijacsa.2019.0100411.
- [38] P. Batra and A. Jatani, "Measurement Based Performance Evaluation of DevOps," in *2020 International Conference on Computational Performance Evaluation, ComPE 2020*, 2020, pp. 757–760. doi: 10.1109/ComPE49325.2020.9200149.
- [39] A. Häkli, D. Taibi, and K. Systa, "Towards Cloud Native Continuous Delivery: An Industrial Experience Report," in *2018 IEEE/ACM International Conference on Utility and Cloud Computing Companion (UCC Companion)*, 2018, pp. 314–320. doi: 10.1109/UCC-Companion.2018.00074.
- [40] D. A. Meedeniya, I. D. Rubasinghe, and I. Perera, "Traceability establishment and visualization of Software Artefacts in DevOps Practice: A survey," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 7, pp. 66–76, 2019, doi: 10.14569/ijacsa.2019.0100711.
- [41] S. M. R. Al Masud, M. Masnun, A. Sultana, A. Sultana, F. Ahmed, and N. Begum, "DevOps Enabled Agile: Combining Agile and DevOps Methodologies for Software Development," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 11, 2022, doi: 10.14569/IJACSA.2022.0131131.

- [42] V. Debroy, S. Miller, and L. Brimble, "Building lean continuous integration and delivery pipelines by applying devops principles: A case study at varidesk," in *ESEC/FSE 2018 - Proceedings of the 2018 26th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, 2018, pp. 851–856. doi: 10.1145/3236024.3275528.
- [43] Suzanna, Sasmoko, F. L. Gaol, and T. Oktavia, "Continuous Software Engineering for Augmented Reality," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 7, 2023, doi: 10.14569/IJACSA.2023.0140719.
- [44] J. Henkel, C. Bird, S. K. Lahiri, and T. Reps, "Learning from, understanding, and supporting devops artifacts for docker," in *Proceedings - International Conference on Software Engineering*, 2020, pp. 38–49. doi: 10.1145/3377811.3380406.
- [45] W. P. Luz, G. Pinto, and R. Bonifácio, "Building a collaborative culture: A grounded theory of well succeeded devops adoption in practice," 2018. doi: 10.1145/3239235.3240299.
- [46] M. Patten, "Questionnaire Research : A Practical Guide," *Quest. Res.*, Oct. 2016, doi: 10.4324/9781315265858.
- [47] K. Finstad, "Response Interpolation and Scale Sensitivity: Evidence Against 5-Point Scales Usability Metric for User Experience View project," *J. Usability Stud.*, vol. 5, no. 3, pp. 104–110, 2010.
- [48] D. Altman, D. Machin, T. Bryant, and M. Gardner, *Statistics with confidence : confidence intervals and statistical guidelines*. Wiley, 2013.
- [49] C. Gerardo, M. Shameem, R. R. Kumar, C. Kumar, B. Chandra, and A. A. Khan, "Prioritizing Challenges of Agile Process in Distributed Software Development Environment Using Analytic Hierarchy Process," *J. Softw. Evol. Process*, vol. 30, no. 11, Nov. 2018, doi: 10.1002/smr.1979.
- [50] E. Albayrak and Y. Erensal, "Using analytic hierarchy process (AHP) to improve human performance: An application of multiple criteria decision making problem: Intelligent Manufacturing Systems: Vision for the Future (Guest Editors: Ercan Öztemel, Cemalettin Kubat and Harun Taşkin)," *J. Intell. Manuf.*, vol. 15, 2004, doi: 10.1023/B:JIMS.0000034112.00652.4c.
- [51] F. T. Bozburu, A. Beskese, and C. Kahraman, "Prioritization of Human Capital Measurement Indicators Using Fuzzy AHP," *Expert Syst. Appl.*, vol. 32, no. 4, pp. 1100–1112, May 2007, doi: 10.1016/j.eswa.2006.02.006.
- [52] S. Garg, P. Pundir, G. Rathee, P. K. Gupta, S. Garg, and S. Ahlawat, "On Continuous Integration / Continuous Delivery for Automated Deployment of Machine Learning Models using MLOps," *arXiv.org. Cornell University Library, arXiv.org PP - Ithaca, Ithaca*, Feb. 07, 2022.
- [53] T. Theo, Uwe, and A. Paris, "A mapping study on documentation in Continuous Software Development," *Inf. Softw. Technol.*, vol. 142, p. 106733, 2022, doi: 10.1016/j.infsof.2021.106733.
- [54] T. Minaoar Hossain, S. Masud, U. Gias, and I. Anindya, "A mixed method study of DevOps challenges," *Inf. Softw. Technol.*, vol. 161, p. 107244, 2023, doi: 10.1016/j.infsof.2023.107244.
- [55] L. Welder Pinheiro, P. Gustavo, and B. Rodrigo, "Adopting DevOps in the real world: A theory, a model, and a case study," *J. Syst. Softw.*, vol. 157, p. 110384, 2019, doi: 10.1016/j.jss.2019.07.083.
- [56] S. Monika, F. Michael, and R. Rudolf, "The pipeline for the continuous development of artificial intelligence models—Current state of research and practice," *J. Syst. Softw.*, vol. 199, p. 111615, 2023, doi: 10.1016/j.jss.2023.111615.
- [57] A. Muhammad Azeem, S. Kari, M. Sajjad, and A. Ahmed, "Toward successful DevSecOps in software development organizations: A decision-making framework," *Inf. Softw. Technol.*, vol. 147, p. 106894, 2022, doi: 10.1016/j.infsof.2022.106894.
- [58] J. Fritsch et al., "Adopting microservices and DevOps in the cyber-physical systems domain: A rapid review and case study," *Softw. - Pract. Exp.*, vol. 53, no. 3, pp. 790–810, 2023, doi: 10.1002/spe.3169.
- [59] I. M. Pereira, T. G. D. S. Carneiro, and E. Figueiredo, "Investigating Continuous Delivery on IoT Systems," 2022. doi: 10.1145/3493244.3493261.
- [60] A. R. Patel and S. Tyagi, "Lightweight Review: Challenges and Benefits of Adopting DevOps," in *Proceedings of 2022 1st International Conference on Informatics, ICI 2022*, 2022, pp. 235–237. doi: 10.1109/ICI53355.2022.9786902.
- [61] M. Rowse and J. Cohen, "A survey of DevOps in the South African software context," *Proc. Annu. Hawaii Int. Conf. Syst. Sci.*, vol. 2020-Janua, pp. 6785–6794, 2021.
- [62] B. Snyder and B. Curtis, "Using Analytics to Guide Improvement during an Agile–DevOps Transformation," *IEEE Softw.*, vol. 35, no. 1, pp. 78–83, 2018, doi: 10.1109/MS.2018.110162910.
- [63] A. Premchand, M. Sandhya, and S. Sankar, "Simplification of application operations using cloud and DevOps," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 13, no. 1, pp. 85–93, 2019, doi: 10.11591/ijeecs.v13.i1.pp85-93.
- [64] G. Pallis, D. Trihinas, A. Tryfonos, and M. Dikaiakos, "DevOps as a Service: Pushing the Boundaries of Microservice Adoption," *IEEE Internet Comput.*, vol. 22, no. 3, pp. 65–71, 2018, doi: 10.1109/MIC.2018.032501519.
- [65] Lewis and R. Jayadi, "IMPLEMENTING CONTINUOUS DELIVERY IN A FINTECH COMPANY: A CASE STUDY," *J. Theor. Appl. Inf. Technol.*, vol. 100, no. 22, pp. 6591–6606, 2022.
- [66] K. Kuusinen et al., "A Large Agile Organization on Its Journey Towards DevOps," 2018 44th Euromicro Conf. Softw. Eng. Adv. Appl., pp. 60–63, Aug. 2018, doi: 10.1109/SEAA.2018.00019.
- [67] S. B. O. G. Caraturan and D. H. Goya, "Major Challenges of Systems-of-Systems with Cloud and DevOps - A Financial Experience Report," in *Proceedings - 2019 IEEE/ACM 7th International Workshop on Software Engineering for Systems-of-Systems and 13th Workshop on Distributed Software Development, Software Ecosystems and Systems-of-Systems, SESoS-WDES 2019*, 2019, pp. 10–17. doi: 10.1109/SESoS/WDES.2019.00010.



APPENDIX A

TABLE XIII. SELECTED PUBLICATION

Authors	Title	Journal/Proceedings	Year	ID
Zhou, X., et al.	A Cross-Company Ethnographic Study on Software Teams for DevOps and Microservices: Organization, Benefits, and Issues	The 44th International Conference on Software Engineering	2022	SP1
Li, H., et al.	Adopting Autonomic Computing Capabilities in Existing Large-Scale Systems	The 40th International Conference on Software Engineering	2018	SP2
Luz, W. P., et al.	Building a Collaborative Culture: A Grounded Theory of Well Succeeded DevOps Adoption in Practice	The 12th ACM/IEEE International Symposium on Empirical Software Engineering	2018	SP3
Debroy, V., et al.	Building Lean Continuous Integration and Delivery Pipelines by Applying DevOps Principles: A Case Study at Varidesk	The 2018 26th ACM Joint Meeting on European Software Engineering Conference	2018	SP4
D'az, J., et al.	DevOps in Practice – An Exploratory Case Study	The 19th International Conference on Agile Software Development	2018	SP5
Henkel, J., et al.	Learning from, Understanding, and Supporting DevOps Artifacts for Docker	The ACM/IEEE 42nd International Conference	2020	SP6
Maroukian, K., et al.	The Link Between Transformational and Servant Leadership in DevOps-Oriented Organizations	The 2020 European Symposium on Software Engineering	2020	SP7
Gupta, R. K., et al.	Challenges in Adopting Continuous Delivery and DevOps in a Globally Distributed Product Team	2019 ACM/IEEE 14th International Conference on Global Software Engineering (ICGSE)	2019	SP8
Bijwe, A. and P. Shankar	Challenges of Adopting DevOps Culture on the Internet of Things Applications - A Solution Model	2022 2nd International Conference on Technological Advancements	2022	SP9
Batra, P. and A. Jatain	Measurement Based Performance Evaluation of DevOps	2020 International Conference on Computational Performance Evaluation	2020	SP10
Häkli, A., et al.	Towards Cloud Native Continuous Delivery: An Industrial Experience Report	2018 IEEE/ACM International Conference on Utility and Cloud Computing Companion	2018	SP11
Lapointe-Boisvert, A., et al.	Towards Modelling Acceptance Tests as a Support for Software Measurement	2021 ACM/IEEE International Conference on Model Driven Engineering Languages	2021	SP12
Shahin, M., et al.	A Qualitative Study of Architectural Design Issues in DevOps	Cornell University Library, arXiv.org	2021	SP13
Sorgalla, J., et al.	Applying Model-Driven Engineering to Stimulate the Adoption of DevOps Processes in Small and Medium-Sized Organizations	Cornell University Library, arXiv.org	2021	SP14
Mauro Lourenço, P., et al.	DevOps ADOPTION: EIGHT EMERGENT PERSPECTIVES	Cornell University Library, arXiv.org	2021	SP15
Trigo, A., et al.	DevOps adoption: Insights from a large European Telco	Cornell University Library, arXiv.org	2022	SP16
Díaz, J., et al.	DevOps in Practice – A preliminary Analysis of two Multinational Companies	Cornell University Library, arXiv.org	2019	SP17
Garg, S., et al.	On Continuous Integration / Continuous Delivery for Automated Deployment of Machine Learning Models using MLOps	Cornell University Library, arXiv.org	2022	SP18
Theo, T., et al.	A mapping study on documentation in Continuous Software Development	Information and Software Technology	2022	SP19
Minaoar Hossain, T., et al.	A mixed method study of DevOps challenges	Information and Software Technology	2023	SP20
Welder Pinheiro, L., et al.	Adopting DevOps in the real world: A theory, a model, and a case study	Journal of Systems and Software	2019	SP21
Monika, S., et al.	The pipeline for the continuous development of artificial intelligence models—Current state of research and practice	Journal of Systems and Software	2023	SP22
Muhammad Azeem, A., et al.	Toward successful DevSecOps in software development organizations: A decision-making framework	Information and Software Technology	2022	SP23
Fritzsch, J., et al.	Adopting microservices and DevOps in the cyber-physical systems domain: A rapid review and case study	Software - Practice and Experience	2023	SP24
Pereira, I. M., et al.	Investigating Continuous Delivery on IoT Systems	Association for Computing Machinery	2022	SP25
Patel, A. R. and S. Tyagi	Lightweight Review: Challenges and Benefits of Adopting DevOps	Proceedings of 2022 1st International Conference on Informatics, ICI 2022	2022	SP26
Rowse, M. and J. Cohen	A survey of DevOps in the South African software context	Proceedings of the Annual Hawaii International Conference on System Sciences	2021	SP27
Snyder, B. and B. Curtis	Using Analytics to Guide Improvement during an Agile-DevOps Transformation	IEEE Software	2018	SP28
Senapathi, M., et al.	DevOps capabilities, practices, and challenges: Insights from a case study	The 22nd International Conference on Evaluation and Assessment	2018	SP29
Premchand, A., et al.	Simplification of application operations using cloud and DevOps	Indonesian Journal of Electrical Engineering and Computer Science	2019	SP30
Pallis, G., et al.	DevOps as a Service: Pushing the Boundaries of Microservice Adoption	IEEE Internet Computing	2018	SP31
Lewis and R. Jayadi	IMPLEMENTING CONTINUOUS DELIVERY IN A FINTECH COMPANY: A CASE STUDY	Journal of Theoretical and Applied Information Technology	2022	SP32
Kuusinen, K., et al.	A Large Agile Organization on Its Journey Towards DevOps	2018 44th Euromicro Conference on Software Engineering and Advanced Applications	2018	SP33
Caraturan, S. B. O. G.	Major Challenges of Systems-of-Systems with Cloud and DevOps - A Financial Experience Report	2019 IEEE/ACM 7th International Workshop on Software Engineering	2019	SP34

APPENDIX B

TABLE XIV. QUALITY RATING OF THE CHOSEN STUDIES

ID	Reference	QA1	QA2	QA3	QA4	QA5	Total	%
SP1	[24]	1	1	1	1	1	5	100%
SP2	[33]	1	0.5	1	1	1	4.5	90%
SP3	[45]	1	1	1	1	1	5	100%
SP4	[42]	1	1	1	1	1	5	100%
SP5	[28]	1	1	1	0.5	1	4.5	90%
SP6	[44]	1	1	1	1	0.5	4.5	90%
SP7	[25]	1	0.5	1	0.5	1	4	80%
SP8	[29]	1	1	1	1	1	5	100%
SP9	[27]	1	0.5	0.5	1	1	4	80%
SP10	[38]	1	0.5	1	1	1	4.5	90%
SP11	[39]	1	0.5	1	0.5	1	4	80%
SP12	[31]	1	0.5	0.5	1	1	4	80%
SP13	[34]	1	1	1	1	1	5	100%
SP14	[35]	1	0.5	0.5	1	1	4	80%
SP15	[32]	1	0.5	1	1	0.5	4	80%
SP16	[3]	1	0.5	1	1	0.5	4	80%
SP17	[26]	1	0.5	1	1	1	4.5	90%
SP18	[52]	1	1	1	1	1	5	100%
SP19	[53]	0.5	0.5	1	1	1	4	80%
SP20	[54]	1	0.5	1	1	1	4.5	90%
SP21	[55]	1	0.5	1	0.5	1	4	80%
SP22	[56]	1	0.5	1	1	1	4.5	90%
SP23	[57]	0.5	0.5	1	0.5	1	3.5	70%
SP24	[58]	0.5	0.5	1	0.5	1	3.5	70%
SP25	[59]	1	0.5	1	1	1	4.5	90%
SP26	[60]	1	0.5	0.5	1	0.5	3.5	70%
SP27	[61]	1	0.5	1	1	1	4.5	90%
SP28	[62]	1	0.5	0.5	1	0.5	3.5	70%
SP29	[11]	1	0.5	1	1	1	4.5	90%
SP30	[63]	0.5	0.5	1	0.5	1	3.5	70%
SP31	[64]	1	0.5	1	1	1	4.5	90%
SP32	[65]	0.5	0.5	1	1	0.5	3.5	70%
SP33	[66]	1	0.5	1	1	1	4.5	90%
SP34	[67]	1	0.5	0.5	1	0.5	3.5	70%

APPENDIX C

TABLE XV. SAMPLE SURVEY QUESTIONNAIRE WAS USED TO CONFIRM THE SECURITY ISSUES WITH DEVOPS

Survey questions to determine the challenges with DevOps adoption in IPHO						
Section A: Personal information of respondents.						
Name						
EmployeeID						
Email address						
How long have you been familiar with or using DevOps?	< 1 year <input type="radio"/>	1 year <input type="radio"/>	2 years <input type="radio"/>	3 years <input type="radio"/>	4 years <input type="radio"/>	5 years <input type="radio"/>
Section B: Challenges related to security in DevOps and their categorization.						
The purpose of this section is to identify Challenges that can have a negative impact on the adoption of DevOps in IPHO. Please provide a rating for each challenge based on your understanding and experience.						
Strongly Agree = 'S-A', Agree = 'A', Neutral = 'N', Disagree = 'D', Strongly Disagree = 'S-D'						
ID	Factors and Categories Identified	S-A	A	N	D	S-D
C1	Does the separation of Developer and Operational teams pose a barrier to implementing DevOps?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
C2	Does ineffective communication channel pose a barrier to implementing DevOps?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
C3	Does the mindset shift from traditional to automated deployment processes pose a barrier to implementing DevOps?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
C4	Does lack of awareness of the benefits of DevOps implementation pose a barrier to its implementation?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
C5	Does resistance to change pose a barrier to implementing DevOps?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
C6	Does adoption of new processes pose a barrier to implementing DevOps?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
C7	Does difficulty in resource allocation pose a barrier to implementing DevOps?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

C8	Does lack of cross-functional leadership pose a barrier to implementing DevOps?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
C9	Does lack of a key leader pose a barrier to implementing DevOps?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
C10	Does lack of performance evaluation with quality metrics pose a barrier to implementing DevOps?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
C11	Does lack of strategic direction and clear definition pose a barrier to implementing DevOps?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
C12	Does lack of management support pose a barrier to implementing DevOps?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
C13	Does lack of examples/guidelines in practice pose a barrier to implementing DevOps?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
C14	Does process complexity pose a barrier to implementing DevOps?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
C15	Does long release cycles pose a barrier to implementing DevOps?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
C16	Does difficulty in learning and disseminating knowledge pose a barrier to implementing DevOps?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
C17	Does lack of staff with good technical skills pose a barrier to implementing DevOps?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
C18	Does weak collaboration pose a barrier to implementing DevOps?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
C19	Does significant effort to transition from manual to automation pose a barrier to implementing DevOps?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
C20	Does difficulty in implementing Automated Testing pose a barrier to implementing DevOps?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
C21	Does difficulty in automating code generation for Infrastructure, Operation, and Test functions pose a barrier to implementing DevOps?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
C22	Does difficulty in implementing good technical documentation pose a barrier to implementing DevOps?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
C23	Does lack of automated testing tools pose a barrier to implementing DevOps?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
C24	Does the use of legacy tools and technologies pose a barrier to implementing DevOps?	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Please give other factors that hinder the implementation of DevOps in IPHO (optional): -----						

# A Bibliometric Analysis of Smart Home Acceptance by the Elderly (2004-2023)

Bo Yuan<sup>1</sup>, Norazlyn Kamal Basha<sup>2</sup>

School of Business and Economics, Universiti Putra Malaysia (UPM), Kuala Lumpur, Malaysia<sup>1,2</sup>

**Abstract**—Both academia and business firmly endorse the notion that a smart home would be the solution to easing the excessive social burden associated with demographic ageing and improving older adults' quality of life by enhancing living independence while encouraging their desire to age in place. This study uses bibliometric analysis to examine the research trends on elderly people's acceptance of smart home. The results are derived from analysis using the VOSviewer software on 257 documents in the Scopus database. The results reveal that: there is an accelerating growth rate for the smart home literature focusing on the elderly's acceptance since 2004; the majority of these studies are journal articles filed in the research area of computer science; the most commonly mentioned keywords include "smart home(s)" and "older adults"; the US has produced the highest number of related works; and the most cited articles are composed by authors across nations with tight collaborations.

**Keywords**—Smart home; acceptance; elderly people; ageing-in-place; bibliometric analysis; VOSviewer

## I. INTRODUCTION

A report from the United Nations [1] revealed a global ageing trend by stating that the worldwide population of people aged 65 and above has grown to 771 million by 2022, and is estimated to reach 1.6 billion by 2050 with an increased proportion of 16 percent from 10 percent (in 2022) of the overall population. Demographic ageing may incur serious social problems, such as overburdened health care systems due to the increased demand for nursing facilities and care services [2],[3]. In addition, elderly people are showing significant willingness to continue living in their own homes [4], [5] for the strong desires of independent living, healthcare and social connection [6]. Meanwhile, the elderly care industry has been driven towards intelligence by the Internet technology sector's quick expansion, so the smart home-based elderly care model has come into being for the reason that smart homes would considerably ease the strain of social health care systems by integrating limited elderly care resources in the market and society to provide elderly care services with wisdom, precision and efficiency [7], and on the other hand, would satisfy the need for ageing-in-place (AIP) by the elderly [2]. Therefore, smart home research from the elderly perspective is of great significance to comprehensively learn how their intentions to accept and utilize smart home technologies are formed.

Smart homes are commonly considered as residences in which devices and appliances equipped with interconnected sensors are placed for the purpose of improving inhabitants' comfort, convenience, safety, security and quality of life [5], [8]. The increasingly expanding consumer interest has led to a

rapid growth in the smart home industry. Tech giants like Google, Amazon, Apple, Huawei, Xiaomi and Samsung are racing to meet the demands for their smart home products and services by competing in this emerging market [9]. As Statista [10] estimated, from US\$115.7 billion in 2022, the global smart home market's revenue is projected to soar to US\$222.9 billion in 2027, a CAGR of 22%. Users can benefit from personalized smart home services through either automation or remote control [11] since data and information are automatically interchanged between devices to make corresponding responses according to changes in surrounding environment or users' customization [12]. With the aid of assistive technologies, the elderly may enjoy a higher quality of life in a "smart home" as evidence has shown that the utilization of nursing homes or care-givers leads to negative impacts on older people, such as feelings of stress, depression and change of habit [3]. In this case, the academia around the world all considered AIP through the adoption of smart home as the optimal solution for ageing societies [13], [14].

The term "smart home" is used in this study to describe homes that provide their residents with superior levels of convenience, security, and comfort, which is actualized by networked devices and appliances with processors and sensors [15]. Elderly people, following the extant research, particularly include people who are aged 60 years old and above [16], [17]. Thus, this study's overarching goal is to identify the current state of research on the topic of elderly citizens' adoption of smart homes, including its driving forces and key players, with the expectation of potentially providing suggestions for future studies in this domain.

The remainder of the study is organized as follows. Based on different research focuses on smart homes, Section II overview the recent literature mainly from the perspective of older adults. Section III describes the employed methodology. The results and findings are presented in Section IV, followed by a detailed discussion on the findings. Section V makes conclusions of the whole study.

## II. LITERATURE REVIEW

### A. From Technology Perspective

There has been a lot of research on many aspects of the smart home for the elderly. One group of researchers looked at how and whether smart home technologies helped the elderly. For example, with the use of Ambient Assisted Living technology, Blackman et al. [18] demonstrated that older individuals' autonomy and quality of life could be improved. Aramendi et al. [19] tried to detect functional health decline in the elderly using in-home behavior data collected by smart

homes, and they concluded that functional health issues are predictable from smart home data, which is important for early intervention in ageing societies. Fritz et al. [20] explored a potentially prominent use of health-monitoring smart homes to provide assistance to older adults with chronic conditions by remotely detecting a range of physical status. Results showed that smart homes helped recognize clinical changes in seniors' health and treatment and medication management. To foresee the likelihood of falls among the elderly, Kulurkar et al. [21] developed a low-cost fall detection system using wearable sensors in a smart home setting, and they ended up with a considerably high accuracy rate of 95.87%.

### B. From User Perspective

Another group of researchers has been looking into smart homes from the viewpoint of the elderly since there is a dearth of literature on the attitudes of older individuals about embracing such technology [22]. According to research by Pignini et al. [23], elderly people value health monitoring systems that employ smart technology for their own protection. Yu et al. [24] clustered Korean seniors based on their residential lifestyles to examine whether or not there are distinct requirements for smart home features, and they found that there was some variation. In a focus-group-based study, Ghorayeb et al. [5] revealed that both users and non-users of smart home monitoring technology considered feasibility, customization and data security of great importance to accept the technology; while users concerned more about utility, and less about privacy intrusion and trust.

### C. Bibliometric and Scientometric Work

Benefit from the rapid growth of information technology and statistics tools, scholars have recently made effort to systematically analyze extant smart home literature through bibliometric data. Choi et al. [25] used a bibliometric approach and analyzed 2339 articles published between 2015 and 2019 from the Scopus database in smart home and Internet of Things (IoT) domain in order to indicate key research trends and knowledge mapping for future studies. Li et al. [26] applied a scientometric analysis with smart home research published from 2000-2021 in the Scopus database to illustrate historical changes, emerging trends, and research clusters. Ohlan and Ohlan [27] comprehensively analyzed bibliometric data published in the Web of Science from 2001 to 2021, and suggested indispensable trends and patterns of smart home research. Another study by Hong et al. [28] applied bibliometric and scientometric analyses with 1408 related articles acquired from the Web of Science database to thoroughly overview smart home features for the elderly.

However, in the existing review studies on smart homes, bibliographic approach was seldom applied, and none of the authors made older adults the central focus of their work. Besides, taking the competence of smart homes in helping ease social stress into consideration in the context of demographic ageing, there is a need for systematic research to comprehensively analyze the research trends on the smart home acceptance literature particularly for the elderly. As far as the authors are aware, no research has been done in the area of Smart Home Acceptance by the Elderly (SHAE). Given that it contributes to substantial insights via bibliometric analysis on

a specific subject, structurally reviewing earlier studies to tease out key research trends becomes crucial for scholars. Therefore, this study aims to investigate current research trends of smart home studies, especially those related to the acceptance by the elderly, applying a bibliometric approach to summarize current directions in this field, prominent venues and authors, research clusters, as well as to identify and suggest prospective research directions. Findings of this study will present relevant stakeholders, such as academics, policymakers, and businesses, with research and working directions for the near future.

## III. METHODOLOGY

### A. Bibliometric Analysis

In the study, published articles in the Scopus database on the global trend of SHAE from 2004 through 2023 (by 25 July) are analyzed using bibliometric analysis. Bibliometric analysis was firstly introduced by Pritchard [29], and has been commonly described as a technique using mathematics to statistically analyze published books, articles, and other communication media so as to provide a broad overview of certain knowledge field [25], [30]. By conducting this technique, academics will gain insight into selected research domains in terms of research trends, significant authors, institutions, publishers, nations, as well as potential research gaps [31].

### B. Data Collection

The Elsevier Scopus database was utilized to collect bibliographic data for this study, which has been recognized as a leading multidisciplinary repository of influential peer-reviewed research in social science fields [25], [32] for the reason that it contains a considerable number of high-quality materials, e.g., 75.5 million files, 24.6 thousand titles, and 194 thousand books [26]. Compared to other commonly used databases for bibliometric analyses (i.e., PubMed and Web of Science), Scopus is advantageous for the reasons that it is the largest database containing multidisciplinary publications, and publications are classified into multiple research areas accordingly [33], which makes it more suitable for mapping based on research areas. And the appropriateness of adopting Scopus database has been particularly acknowledged by scholars in the smart home domain [25], [26]. Thus, from the time the first publication on SHAE appeared in the Scopus database until this study was composed, all publications on the subject of the smart home acceptance by the elderly were included (25 July, 2023).

The objective of this research is to examine, from a user's point of view, the relationships between and clusters of related studies on smart homes. Therefore, the query (TITLE-ABS-KEY (smart AND home) AND (adoption OR acceptance)) was set as the search task, which produced a number of 1421 documents. Based on the search results, the query was then refined as (TITLE-ABS-KEY (smart AND home) AND (adoption OR acceptance) AND (elderly OR older)) to focus on the predetermined target age-group, and this resulted in a reduced number of 257 documents. All of the generated documents from the refined searching are retained for a thorough analysis and understanding of the research trends.

The data was collected on 25 July, 2023. The search result includes conference papers, articles, book chapters, reviews, conference reviews, books, notes, letters, and short survey. Data were exported in “CSV” format to be compatible with VOSviewer, the software employed for data analysis.

IV. RESULTS

A. Publication Trend

Table I summarized the publication trend by year since the first article on the elderly acceptance in smart home industry published in 2004. It is clear that not all research articles have been cited. Except for the first six years (2004 – 2009), the proportions of cited publications over total yearly publications never reached 100% for the following years onwards. Overall, the average proportion was 77.14% with a standard deviation of 13.10 from 2010 to 2023. The number of citations received also varied a lot. Documents published in 2018 seem to be most influential since they gained the highest number of citations of 855. Besides, publications in the years of 2008, 2014, 2016 and 2019 also had significant impact for the academia with citations over 460, respectively.

TABLE I. PUBLICATION TREND

Year	Number of Publications	Cited Publications	Proportion	Total Citations
2004	2	2	100%	67
2005	0	0	-	0
2006	1	1	100%	1
2007	2	2	100%	50
2008	5	5	100%	503
2009	3	3	100%	40
2010	6	5	83%	69
2011	10	8	80%	256
2012	13	11	85%	204
2013	9	6	67%	339
2014	13	10	77%	520
2015	10	9	90%	202
2016	13	12	92%	505
2017	17	13	76%	341
2018	22	15	68%	874
2019	20	16	80%	480
2020	23	21	91%	285
2021	30	26	87%	359
2022	41	26	63%	92
2023	17	6	35%	17
<b>Total</b>	<b>257</b>	<b>197</b>		<b>5204</b>

The publication trend can be seen in three stages in terms of research productivity, which is shown in Fig. 1. The first publication appeared in 2004, and was composed by Barlow and Venables [34], entitled “Will technological innovation create the true lifetime home”. For the first stage (2004-2009), no more than five documents were published each year. With an average of 2.2 publications per year, 13 publications (5.06%) in total were produced throughout this time period. In the second stage (2010-2017), a steady growth emerged, and yearly publications ranged from 6 to 17. Overall, 91 publications (35.41%) were produced with an annual average of 11.4. Starting from 2011, a threshold of minimum ten articles (except for 2013, 9 articles) were achieved every year. The third stage extends throughout the most recent six years (2018-2023). Productivity increased considerably as

researchers shifted their focus to the positive or negative effects of smart homes on the elderly. The lower bound for annual publications in this phase increased to 20 from the previous level of 6 with an exception of 17 publications in 2023, which is just the middle of the year. Notably, scholars made the most publications of 41 solely in 2022, and the total publications since 2020 (111) has way exceeded the sum of the entire second stage (91).

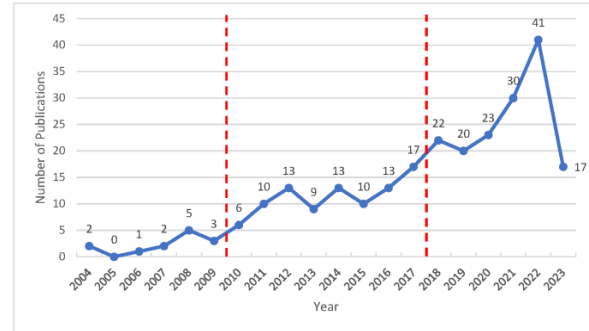


Fig. 1. Publication trend.

B. Types of Documents & Sources

By analyzing the data obtained from the Scopus database, seven types of documents were identified, i.e., article, conference paper, conference review, review, book chapter, book, and note. As displayed in Table II, the proportion of “article” (43.58%) is the highest, followed by “conference paper” (34.24%). While other document types account for less than 10% respectively, and there are only one “book” and one “note” published only within the investigated time range.

TABLE II. SUMMARY OF DOCUMENT TYPES

Document Types	Number of Publications	Proportion
Article	112	43.58%
Conference Paper	88	34.24%
Conference Review	22	8.56%
Review	20	7.78%
Book Chapter	13	5.06%
Book	1	0.39%
Note	1	0.39%
<b>Total</b>	<b>257</b>	<b>100%</b>

From Table III, all publications related to the elderly people's openness to smart homes fall into four source types. More than half of the total documents were published in “journal” (52.14%). Not much differences were found between “conference proceeding” (24.12%) and “book series” (19.07%), while “book” seems to be the least interested by researchers since only 12 have been published throughout 19 years since 2004.

TABLE III. SUMMARY OF SOURCE TYPES

Source Types	Number of Publications	Proportion
Journal	134	52.14%
Conference Proceeding	62	24.12%
Book Series	49	19.07%
Book	12	4.67%
<b>Total</b>	<b>257</b>	<b>100%</b>

C. Subject Areas

In all, 21 research areas were addressed by the 257 documents. The distribution of the top ten subject areas related to older adults' adoption of smart homes over a five-year period is shown in Table IV. The density of publications is shown by the red backdrop. The more papers that emerged within a certain year period, the deeper is the shade of red. "Computer Science" (140 publications) has the highest number of articles that are relevant; it is followed in relevance by "Medicine" (80 publications), "Engineering" (70 publications), "Mathematics" (46 publications), and "Social Sciences" (40 publications). It seems that the majority of the studies examining the benefits of smart homes for the elderly focus on their technical and medical applications. In contrast to recent increase in "Biochemistry", "Genetics", and "Molecular Biology", subject areas such as "Engineering", "Social Sciences", and "Physics and Astronomy" exhibit constant growth. "Health Professions" shows a brief decline followed by a minor recent rise.

TABLE IV. DISTRIBUTION OF TOP TEN SUBJECT AREAS IN FIVE-YEAR PERIOD

Table with 6 columns: Research Area, 2004-2008, 2009-2013, 2014-2018, 2019-2023, Total. Rows include Computer Science, Medicine, Engineering, Mathematics, Social Sciences, Nursing, Biochemistry, Genetics and Molecular Biology, Health Professions, Physics and Astronomy, Business, Management and Accounting.

D. Source Title

Research on the SHAE is covered by 160 publication sources. With a minimum of three publications, Table V lists the publications sources with the most activity. The greatest numbers of studies relating to computer science fields are presented in the form of conference proceedings, which are included in the source of "Lecture Notes in Computer Science" (31, 12.1%). When it comes to producing high-quality conference materials, "ACM International Conference Proceeding Series" is another top pick, coming in at number four on the list of all source titles. "Gerontechnology", "Communications In Computer And Information Science" and "Jmir Aging" are the three most prolific journals, with 7, 6, and 5 publications, respectively. Each of the remaining sources contributes less than 2%.

E. Author Keywords

Understanding the distribution and links between the primary research themes on the SHAE was facilitated by the use of co-occurrence analysis, which may probe the internal relationships of a given academic subject [35]. Based on

"keywords" of bibliographic data, a map is produced using the VOSviewer software co-occurrence analysis. Setting the minimum occurrence number to five resulted in a clear keyword network visualization after a number of optimization processes that varied the minimum number of keyword occurrences. The results created a graph with 32 keywords based on five clusters, each with a different focus (see Fig. 2). Cluster 1 (red) focuses on assistive technology for older adults, Cluster 2 (green) emphasizes smart home functionality, Cluster 3 (blue) highlights the Internet of Things and healthcare, Cluster 4 (olive) emphasizes the acceptance of smart home technology, and Cluster 5 (purple) focuses on older people and gerontechnology. Generally, the most common keywords related to older adults and smart homes are included in Cluster 1, while Cluster 4 is mainly about aging and disease.

TABLE V. PRODUCTIVE PUBLICATION SOURCES

Table with 3 columns: Source Title, Publications, Proportion. Rows include Lecture Notes In Computer Science, Gerontechnology, Communications In Computer And Information Science, ACM International Conference Proceeding Series, IEEE Access, Jmir Aging, International Journal Of Environmental Research And Public Health, Gerontology, Handbook Of Smart Homes Health Care And Well Being, International Journal Of Medical Informatics, Journal Of Medical Internet Research, Personal And Ubiquitous Computing, Procedia Computer Science, Proceedings Of The ACM On Human Computer Interaction, Sensors, Universal Access In The Information Society, Others.

In particular, regarding the topic of ageing-in-place and the use of smart technology by the elderly, the most often occurring terms are concentrated in Cluster 1, which has a total link strength of 168. The keyword "smart home" dominates Cluster 2 and has the greatest frequency of occurrence and total link strength of any other term generated. Cluster 3 is somewhat related to Cluster 2 and focuses on the application of IoT technology in the field of geriatric care due to the tight relationship between the technical features of smart homes and IoT. Ageing and the acceptance of smart home technologies are the two topics majorly covered under Cluster 4. With the lowest total link strength of 84 among all five clusters, Cluster 5 focuses primarily on investigating elderly-oriented technologies.

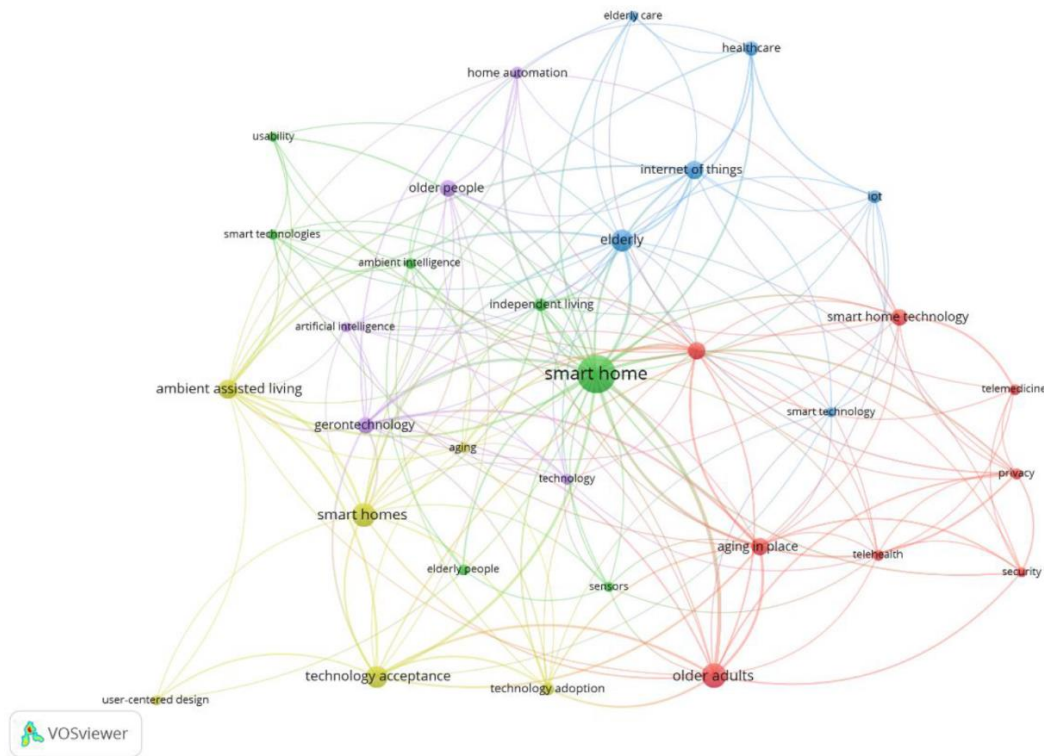


Fig. 2. Co-occurrence network by author keywords.

Table VI lists all keywords generated from related publications in a cluster view. All clusters contain keywords relating to both the elderly (e.g., “older adults”, “elderly people”, “elderly”, “ageing”, “older people”) and technological concepts (e.g., “assistive technology”, “ambient intelligence”, “internet of things”, “ambient assisted living”, “gerontechnology”). However, only the fourth cluster specifically concerns about individuals’ willingness to accept or adopt smart home technology. This is consistent with the viewpoint that there is a lack of studies regarding smart homes from the perspective of users’ perceptions of embracing smart home technologies [12], especially for the elderly people [5].

TABLE VI. KEYWORDS BY CLUSTERS

Cluster	Keyword	Occurrence	Total Link Strength
Cluster 1 (red)	older adults	26	39
	aging in place	14	37
	assistive technology	14	29
	smart home technology	12	16
	privacy	6	17
	telehealth	6	10
	telemedicine	6	8
	security	5	12
Cluster 2 (green)	smart home	58	86
	independent living	8	17
	ambient intelligence	6	13
	elderly people	6	5
	smart technologies	5	10
	sensors	5	9
	usability	5	7
Cluster 3	elderly	21	35

Cluster	Keyword	Occurrence	Total Link Strength
(blue)	internet of things	15	24
	healthcare	10	13
	iot	8	10
	smart technology	5	8
	elderly care	5	7
Cluster 4 (olive)	smart homes	24	33
	technology acceptance	21	33
	ambient assisted living	16	33
	technology adoption	8	18
	ageing	5	12
Cluster 5 (purple)	user-centered design	5	5
	gerontechnology	12	34
	older people	13	15
	home automation	7	14
	technology	6	10
	artificial intelligence	5	11

#### F. Citation Analysis

The top ten most influential papers in the field of SHAE as determined by citation analysis are listed in Table VII of the findings. The article with the highest total citations addressed elderly people’s adoption of voice-activated smart homes and proposed corresponding benefits and concerns [36]. After this article appeared in 2013, it has attracted 305 citations. The second-highest cited work developed a conceptual model to describe the influencing elements of older individuals’ usage of technology that support ageing-in-place [37]. A total of 276 citations have been made to this article since its publication in 2016. Even though over 42% of the publications were merely from the top five most prolific countries, it is worth noting that



leading studies in the field of SHAE were authored by researchers from a variety of nations.

TABLE VII. TOP TEN CITED ARTICLES

Author(s)	Article Title	Year	Source Title	Citations
Portet, F., Vacher, M., Golanski, C., Roux, C., Meillon, B.	Design and evaluation of a smart home voice interface for the elderly: acceptability and objection aspects [36]	2013	Personal and Ubiquitous Computing	305
Peek, S. T. M., Luijckx, K. G., Rijnaard, M. D., Nieboer, M. E., van der Voort, C. S., Aarts, S., van Hoof, J., Vrijhoef, H. J. M., Wouters, E. J. M.	Older Adults' Reasons for Using Technology while Aging in Place [37]	2016	Gerontology	276
Robinson, H., MacDonald, B., Broadbent, E.	The Role of Healthcare Robots for Older People at Home: A Review [38]	2014	International Journal of Social Robotics	261
Pal, D., Funilkul, S., Charoenkitkarn, N., Kanthamanon, P.	Internet-of-Things and Smart Homes for Elderly Healthcare: An End User Perspective [39]	2018	IEEE Access	185
Bansal, P., Kockelman, K. M.	Are we ready to embrace connected and self-driving vehicles? A case study of Texans [40]	2016	Transportation	162
Mital, M., Chang, V., Choudhary, P., Papa, A., Pani, A. K.	Adoption of Internet of Things in India: A test of competing models using a structured equation modelling approach [41]	2018	Technological Forecasting and Social Change	157
Shin, J., Park, Y., Lee, D.	Who will be smart home users? An analysis of adoption and diffusion of smart homes [15]	2018	Technological Forecasting and Social Change	152
Courtney, K., Demiris, G., Rantz, M., Skubic, M.	Needing smart home technologies: the perspectives of older adults in continuing care retirement communities [42]	2008	Journal of Innovation in Health Informatics	144
Demiris, G., Oliver, D. P.,	Findings from a participatory	2008	Technology and Health	134

Author(s)	Article Title	Year	Source Title	Citations
Dickey, G., Skubic, M., Rantz, M.	evaluation of a smart home application for older adults [43]		Care	
Courtney, K.L.	Privacy and Senior Willingness to Adopt Smart Home Information Technology in Residential Care Facilities [44]	2008	Methods of Information in Medicine	126

### G. Co-Authorship Analysis by Countries

The analysis of international co-authorship networks covered only nations with at least five total publications. Only 18 of the 43 countries satisfied the criteria. There are six European countries, two American, one Asian, and one Oceanian among the top ten most productive countries listed in Table VIII. The US, Germany, and the UK are the top three nations. In terms of publications, the US is well ahead of any other country in the SHAE field. C/P value, which measures the average citation count in a single publication, places the Netherlands, France, and Austria among the leaders in this field despite their relatively low total number of publications.

TABLE VIII. TOP TEN PRODUCTIVE COUNTRIES

Country	Number of Publications (P)	Citations (C)	C/P	Total Link Strength
US	46	1308	28.43	7
Germany	27	434	16.07	11
UK	27	538	19.93	9
Italy	22	555	25.23	12
Canada	19	213	11.21	7
China	15	198	13.20	7
Netherlands	15	680	45.33	11
Australia	13	158	12.15	3
France	11	418	38.00	6
Austria	8	300	37.50	7

Fig. 3 show the national network. The size of the node expands as more articles are published in a single country, and the connecting lines between nodes represent the closeness of the collaboration between the two nations; a thicker line indicates greater cooperation. This network consists of 35 links and 4 clusters. Within each of their different clusters, the US, Germany, UK, and China all contribute significantly to publications, and with the exception of Germany, they are all tightly connected. Since Italy, Germany, and Netherlands have the top three total link strengths (12 for Italy, 11 for Germany, and 11 for Netherlands) among the 18 countries, there are strong lines connecting them, showing tight collaborations. And, in terms of publications, they also ranked highly.

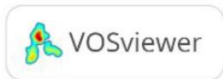
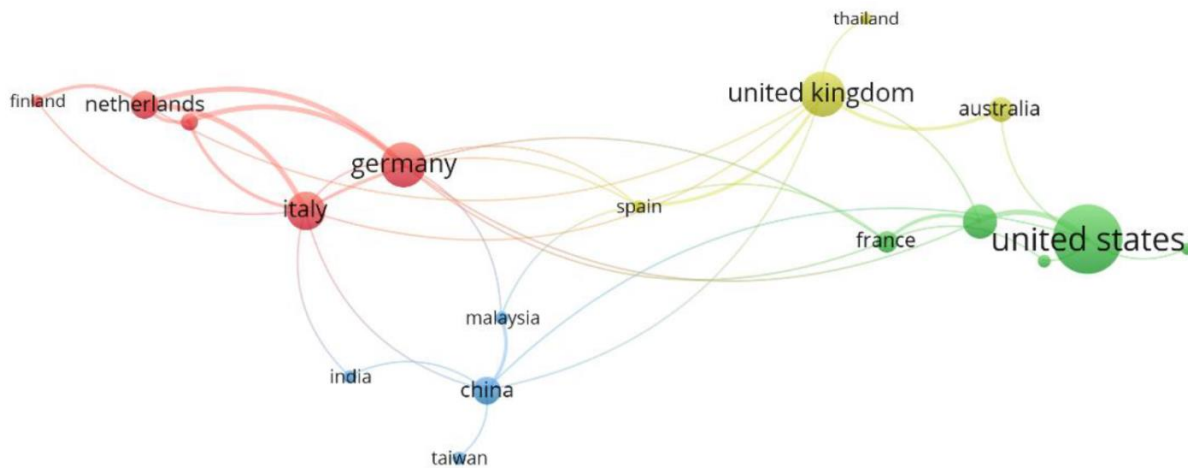


Fig. 3. Co-authorship network by countries.

#### H. Co-Authorship Analysis by Authors

Modern research has entered the age of big science, which emphasizes the need of scientific cooperation due to increased comprehensiveness and complexity [45], [46]. As a result of the complementarity in knowledge and intelligence, academics tend to form solid connections and network blocks [47], and sustain persistent collaborations [48]. Table IX provides a summary of the top ten prolific writers on the SHAE by concentrating on writers who have coauthored at least two papers and been referenced in no less than four other works. Demiris, G. has published the most articles with the highest citations collected; however, his C/P figure only ranked the 8th. Aarts, S., Peek, S.T.M. and Wouters, E.J.M. are closely collaborating with the same publications of 3, and citations of 325. The same pattern has been detected between Portet, F. and Vacher, M.

The scholars who have co-authored with the most other authors is analyzed by the co-authorship network analysis. Co-authorship relationships revealed in research on seniors' acceptance of smart homes are shown in Fig. 4. The constructed network has 63 authors that satisfy the aforementioned criteria with 109 links and a total link strength of 197. The authors are divided into 21 groups, each of which is indicated by a distinct color. The size of each node in the network, which represents an individual author, reflects its relative importance in the network. The findings reveal that Aarts S., Peek S.T.M., and Wouters E.J.M. are top three biggest author nodes and that they are all members of the same cluster (green), which has the greatest total link strength of

134. And academics from the Netherlands predominate, with many members having ties to the same institution, such as Maastricht University and Tilburg University. The interest areas of this group include older adults' technology acceptance, ageing-in-place and healthcare. For example, Peek S.T.M. "Older Adults' Reasons for Using Technology while Aging in Place" [37] in 2016, the most highly collaborative article (9 authors), and "Factors influencing acceptance of technology for aging in place: a systematic review" [49] in 2014, the most cited article (1044 times). In another group (red), scholars are from various countries (e.g., Austria, UK, Netherlands and US), and this group specializes in social applications of intelligent technology.

TABLE IX. TOP TEN PROLIFIC AUTHORS

Author	Publications (P)	Citations (C)	C/P	Total Link Strength
Demiris, G.	9	489	54.33	15
Ziefle, M.	9	174	19.33	10
Brauner, P.	4	42	10.5	7
Funilkul, S.	4	310	77.5	8
Pal, D.	4	310	77.5	8
Aarts, S.	3	325	108.33	17
Peek, S.T.M.	3	325	108.33	17
Portet, F.	3	324	108	9
Vacher, M.	3	324	108	9
Wouters, E.J.M.	3	325	108.33	17

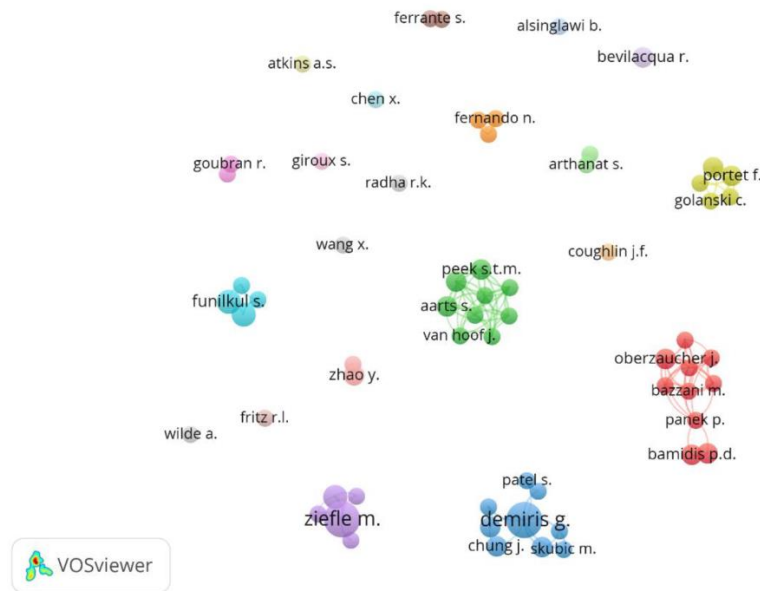


Fig. 4. Co-authorship network by authors.

### I. Bibliographic Coupling

Bibliographic coupling happens when a third source is often cited in two distinct research publications. A visualized bibliographic coupling map at the country level (except for Taiwan) is shown in Fig. 5 with five clusters denoted by different colors, 100 linkages, and a total link strength of 3188. Totally, 18 countries and district that contributed at least five documents are covered.

In terms of published documents, the US, UK, Germany, and Italy are the most substantial nodes. The linking lines indicate the coupling relationship between countries/district

and suggest that nations from various parts of the world are linked by similar patterns of citing reference in their research articles. There are a number of thick lines between countries/district from the same or different clusters, for example, China and Malaysia (the highest link strength of 265), UK and Netherlands (link strength of 199), Germany and Malaysia (link strength of 168), which reveal that there exist collaborations between these connected countries/district. In sum, the investigated data set shows a significant bibliographic connection between the countries and district in different parts of the world.

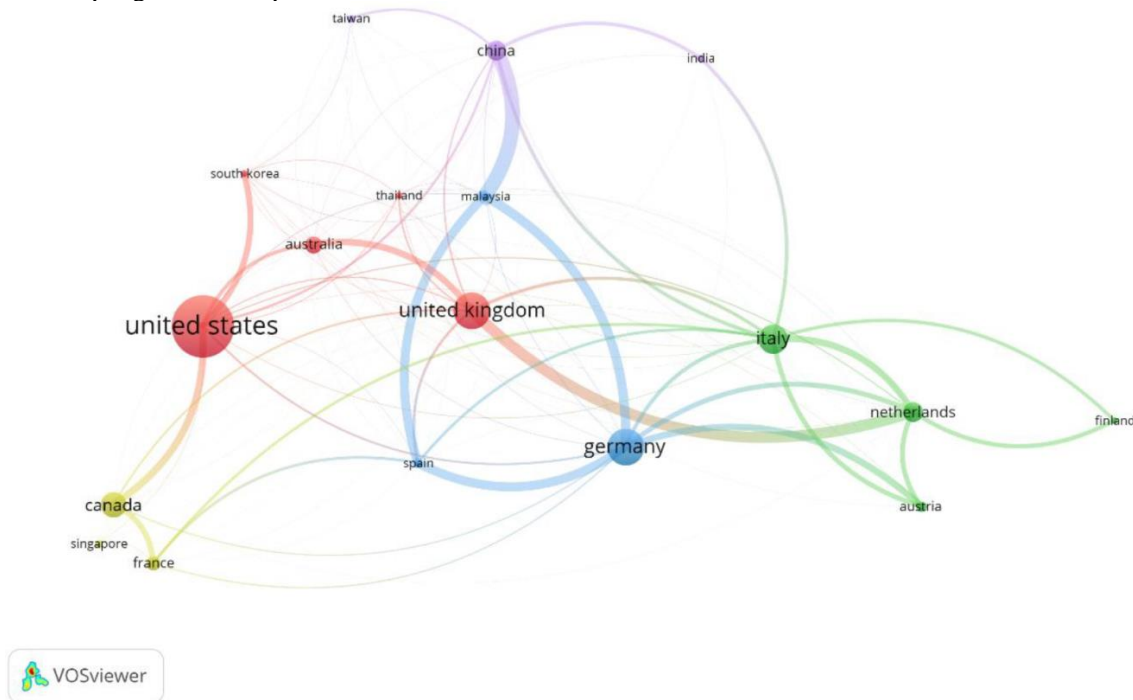


Fig. 5. Bibliographic coupling of countries/district.

## V. DISCUSSION

Due to the worldwide increasingly growing number of ageing population and the associated social issues, research about smart home focusing on the elderly population has been rapidly expanding. However, it has become challenging for academics and practitioners to have a thorough understanding of SHAE due to the enormous rise in the number of related publications. Thus, based on a number of research facets, this study offers a bibliometric analysis of studies on SHAE. The results are discussed in this section regarding each of these facets.

### A. Growing Publication Trend

Since its first presentation, SHAE literature has seen a rising tide of publications, indicating growing interest in the field. Rapid development of the Internet of Things (IoT) has spawned growing practical applications across a wide range of industries, which may account for the uptick in research activity [50]. Additionally, in response to the worldwide challenges (e.g., ageing progress, resources shortage, COVID-19 pandemic), many in academia, business, and government see smart homes as a way to help solve these societal problems [26].

### B. Lack of Interdisciplinary Collaboration

Over half of the current SHAE research may be categorized within the domains of "Computer Science," "Medicine," and "Engineering," according to a breakdown of the most important research topics based on the categories offered by Scopus. Moreover, the majority (13) of the 15 cited published sources focused on either computer-related topics or medically-related applications. Even though knowledge from some cross-disciplinary domains (e.g., psychology, arts and humanities, biology) were referred to, the frequencies were fairly low. Thus, researchers commonly made suggestions and appeal for more attempt at conducting interdisciplinary work [26], [51],[52]. Possible reasons for the currently less convergent research pattern might be a consequence of the competition between tech giants and with other industries [53], which limits the access of specialized knowledge, labor and services [54], then in turn hindered the interdisciplinary collaborations.

### C. Little Emphasis on Elderly Users

As the rising proportion of the ageing population and the trend of AIP while applying smart home technologies [2], More studies on how well smart homes are accepted by the elderly are urgently needed. The results of keywords analysis show that technology-related terms appeared in all keyword clusters, such as "assistive technology", "ambient intelligence", "internet of things", "ambient assisted living", and "gerontechnology". This indicates that the main topics covered in SHAE research are centered on technology attributes and advances, which has been presented in extant research [26].

Conversely, user-oriented keywords such as "technology acceptance", "technology adoption" and "user-centered design" appeared 19, 8 and 5 times respectively, accounting for only 9 percent totally, which agrees with the argument that few research has considered the special needs and using experience of older people in the development and utilization of smart home technologies [55], [56]. Concerns about home and

personal network security have been heightened as a consequence of the intensive adoption of work-from-home strategies due to the COVID-19 pandemic [57], and thus, smart home research extensively put focus on enhancing technological features to ease the concerns relating to possible risks [58]. So, technological advance is currently dominating the research interest and prevalence of existing smart home studies with little emphasis on users' perceptions of the challenges and benefits while using smart home technologies [12], [28]. Furthermore, the ways in which technology is changing the life of the elderly has been surprisingly neglected in research on AIP in smart home settings [59]. Therefore, more elderly user-oriented research on SHAE is needed in order to offer more precise insights into this industry from the demand perspective.

## VI. CONCLUSION

In order to comprehend research publishing trends, document and source types, subject areas, popular source titles, research keyword clusters, notable authors, co-authorship networks, and bibliographic coupling patterns, this study performed bibliometric analyses on the SHAE research since 2004 till 25 July, 2023. By scanning the Scopus database for keywords related to the smart home acceptance by the elderly, the study conducted all analyses with a dataset including 249 documents using the VOSviewer software. Based on the results, this study draws conclusions that the elderly-friendly smart home has dramatically gained academic interest on a global scale in terms of the rapid rise in publications during the last five years; collaborations across disciplines is needed for further research progress in SHAE; the "Lecture Notes In Computer Science" appears to be the most common research venue; technology and health related keywords are all crucial terms other than "smart home(s)" and "older adults"; there exists some global scientific collaboration; and the coincidence of citing preferences at the country level is proved; while the co-authorship network between individual authors is not yet well-established. Thus, researchers should put more attention on learning elderly users' acceptance toward smart homes that are considered promising solution to ageing issues.

Implications of the study are as followed. Firstly, this is the first study, to the authors' knowledge, that employs a comprehensive bibliometric approach to analyze the global trends in research on the acceptance of smart homes by the elderly population. Quantitative information on essential knowledge is provided by the results of the study, which may be utilized to identify research gaps and to plan for future research. Secondly, for policymakers relating to smart homes, more evidence-driven decisions on resource allocation and investment priority could be made according to the findings of this study. Thirdly, insights gained from this study will help develop both national and international strategies for the smart home market.

The results should not be taken as definitive because of the study's limitations. Firstly, the scope of the study could be expanded by including more research databases (e.g., Web of Science) that index additional significant research publications. Secondly, clusters presented in the study are not absolute since the numbers are specified by the criteria determined by the

authors. Thirdly, the VOSviewer software may incur imprecise and complex results due to the issue of regarding single and plural forms of the same term as different words (e.g., “smart home” and “smart homes”).

Therefore, future research would benefit greatly from either an improved version of the VOSviewer software or the inclusion of additional tools for network analysis and visualization. Moreover, as a consequence of the COVID-19 pandemic and the accompanying restrictions on working and living at homes, people have been forced to engage in substantial and mandatory technological learning behaviors [60]. Thus, possible future study directions include accounting for the pandemic's impact on the adoption of smart home technologies among the elderly.

#### ACKNOWLEDGMENT

The authors of this study are appreciative to the reviewers for their insightful feedback and valuable recommendations that will help make this article even better.

#### REFERENCES

- [1] Department of Economic and Social Affairs, Population Division (2022). World Population Prospects 2022. In United Nation (Issue 9). [www.un.org/development/desa/pd/](http://www.un.org/development/desa/pd/).
- [2] Creaney, R., Reid, L., & Currie, M. (2021). The contribution of healthcare smart homes to older peoples' wellbeing: A new conceptual framework. *Wellbeing, Space and Society*, 2(March), 100031. <https://doi.org/10.1016/j.wss.2021.100031>
- [3] Llumiguano, H., Espinosa, M., Jiménez, S., Fernandez-Bermejo, J., del Toro, X., & López, J. C. (2022). An open and private-by-design active and Healthy Ageing Smart Home Platform. *Procedia Computer Science*, 207, 13–23. <https://doi.org/10.1016/j.procs.2022.09.033>
- [4] Fritz, R. L., & Dermody, G. (2019). A nurse-driven method for developing artificial intelligence in “smart” homes for aging-in-place. *Nursing Outlook*, 67(2), 140–153. <https://doi.org/10.1016/j.outlook.2018.11.004>
- [5] Ghorayeb, A., Comber, R., & Goberman-Hill, R. (2021). Older adults' perspectives of Smart Home Technology: Are we developing the technology that older people want? *International Journal of Human-Computer Studies*, 147, 102571. <https://doi.org/10.1016/j.ijhcs.2020.102571>
- [6] Wiles, J. L., Leibling, A., Guberman, N., Reeve, J., & Allen, R. E. (2012). The meaning of "aging in place" to older people. *The Gerontologist*, 52(3), 357–366. <https://doi.org/10.1093/geront/gnr098>
- [7] Wang, W., & Yan, J. (2023). Smart Home-based Elderly Care: the Design of All Service Elements under O2O Mode. *Times of Economy & Trade*, 2023(1), 144–149. <https://doi.org/10.19463/j.cnki.sdjm.2023.01.032>
- [8] Liu, P., Li, G., Jiang, S., Liu, Y., Leng, M., Zhao, J., Wang, S., Meng, X., Shang, B., Chen, L., & Huang, S. H. (2019). The effect of smart homes on older adults with chronic conditions: A systematic review and meta-analysis. *Geriatric Nursing*, 40(5), 522 - 530. <https://doi.org/10.1016/j.gerinurse.2019.03.016>
- [9] Ferreira, L., Oliveira, T., & Neves, C. (2023). Consumer's intention to use and recommend Smart Home Technologies: The role of environmental awareness. *Energy*, 263, 125814. <https://doi.org/10.1016/j.energy.2022.125814>
- [10] Statista, Smart Home – Market Data & Forecast 2022, 2022, [Online; accessed: March 2, 2023]. URL <https://www.statista.com/study/42112/smart-home-report/>.
- [11] Pal, D., Funilkul, S., Vanijja, V., & Papisratorn, B. (2018). Analyzing the elderly users' adoption of smart-home services. *IEEE Access*, 6, 51238–51252. <https://doi.org/10.1109/access.2018.2869599>
- [12] Marikyan, D., Papagiannidis, S., & Alamanos, E. (2019). A systematic review of the Smart Home Literature: A user perspective. *Technological Forecasting and Social Change*, 138, 139–154. <https://doi.org/10.1016/j.techfore.2018.08.015>
- [13] Moyle, W., Murfield, J., & Lion, K. (2021). The effectiveness of smart home technologies to support the health outcomes of community-dwelling older adults living with dementia: A scoping review. *International Journal of Medical Informatics*, 153, 104513. <https://doi.org/10.1016/j.ijmedinf.2021.104513>
- [14] Mihailidis, A., Cockburn, A., Longley, C., & Boger, J. (2008). The acceptability of home monitoring technology among community-dwelling older adults and Baby Boomers. *Assistive Technology*, 20(1), 1–12. <https://doi.org/10.1080/10400435.2008.10131927>
- [15] Shin, J., Park, Y., & Lee, D. (2018). Who will be smart home users? an analysis of adoption and diffusion of smart homes. *Technological Forecasting and Social Change*, 134, 246–253. <https://doi.org/10.1016/j.techfore.2018.06.029>
- [16] Liu, L., Stroulia, E., Nikolaidis, I., Miguel-Cruz, A., & Rios Rincon, A. (2016). Smart Homes and Home Health Monitoring Technologies for Older Adults: A systematic review. *International Journal of Medical Informatics*, 91, 44–59. <https://doi.org/10.1016/j.ijmedinf.2016.04.007>
- [17] Ma, B., Yang, J., Wong, F. K., Wong, A. K., Ma, T., Meng, J., Zhao, Y., Wang, Y., & Lu, Q. (2023). Artificial Intelligence in elderly healthcare: A scoping review. *Ageing Research Reviews*, 83, 101808. <https://doi.org/10.1016/j.arr.2022.101808>
- [18] Blackman, S., Matlo, C., Bobrovitskiy, C., Waldoch, A., Fang, M. L., Jackson, P., Mihailidis, A., Nygård, L., Astell, A., & Sixsmith, A. (2016). Ambient Assisted Living Technologies for aging well: A scoping review. *Journal of Intelligent Systems*, 25(1), 55–69. <https://doi.org/10.1515/jisys-2014-0136>
- [19] Alberdi Aramendi, A., Weakley, A., Aztiria Goenaga, A., Schmitter-Edgcombe, M., & Cook, D. J. (2018). Automatic assessment of functional health decline in older adults based on Smart Home Data. *Journal of Biomedical Informatics*, 81, 119–130. <https://doi.org/10.1016/j.jbi.2018.03.009>
- [20] Fritz, R., Wuestney, K., Dermody, G., & Cook, D. J. (2022). Nurse-in-the-loop smart home detection of health events associated with diagnosed chronic conditions: A case-event series. *International Journal of Nursing Studies Advances*, 4, 100081. <https://doi.org/10.1016/j.ijnsa.2022.100081>
- [21] Kulurkar, P., Dixit, C. kumar, Bharathi, V. C., Monikavishnuvarthini, A., Dhakne, A., & Preethi, P. (2023). AI based elderly fall prediction system using wearable sensors: A smart home-care technology with IOT. *Measurement: Sensors*, 25, 100614. <https://doi.org/10.1016/j.measen.2022.100614>
- [22] Lee, L. N., & Kim, M. J. (2020). A critical review of smart residential environments for older adults with a focus on pleasurable experience. *Frontiers in Psychology*, 10. <https://doi.org/10.3389/fpsyg.2019.03080>
- [23] Pignini, L., Bovi, G., Panzarino, C., Gower, V., Ferratini, M., Andreoni, G., Sassi, R., Rivolta, M. W., & Ferrarin, M. (2017). Pilot test of a new personal health system integrating environmental and wearable sensors for telemonitoring and care of elderly people at Home (Smarta Project). *Gerontology*, 63(3), 281–286. <https://doi.org/10.1159/000455168>
- [24] Yu, J., de Antonio, A., & Villalba-Mora, E. (2020). Older adult segmentation according to residentially-based lifestyles and analysis of their needs for smart home functions. *International Journal of Environmental Research and Public Health*, 17(22), 8492. <https://doi.org/10.3390/ijerph17228492>
- [25] Choi, W., Kim, J., Lee, S. E., & Park, E. (2021). Smart Home and internet of things: A bibliometric study. *Journal of Cleaner Production*, 301, 126908. <https://doi.org/10.1016/j.jclepro.2021.126908>
- [26] Li, W., Yigitcanlar, T., Liu, A., & Erol, I. (2022). Mapping two decades of Smart Home Research: A Systematic Scientometric analysis. *Technological Forecasting and Social Change*, 179, 121676. <https://doi.org/10.1016/j.techfore.2022.121676>
- [27] Ohlan, R., & Ohlan, A. (2022). A comprehensive bibliometric analysis and visualization of Smart Home Research. *Technological Forecasting and Social Change*, 184, 121975. <https://doi.org/10.1016/j.techfore.2022.121975>
- [28] Hong, Y.-K., Wang, Z.-Y., & Cho, J. Y. (2022). Global research trends on smart homes for older adults: Bibliometric and Scientometric

- analyses. *International Journal of Environmental Research and Public Health*, 19(22), 14821. <https://doi.org/10.3390/ijerph192214821>
- [29] Pritchard, A. (1969). *Statistical bibliography or bibliometrics*. *Journal of Documentation*, 25, 348-349
- [30] Ifitkhar, P. M., Ali, F., Faisaluddin, M., Khayyat, A., De Gouvía De Sa, M., & Rao, T. (2019). A bibliometric analysis of the top 30 most-cited articles in gestational diabetes mellitus literature (1946-2019). *Cureus*. <https://doi.org/10.7759/cureus.4131>
- [31] Merigó, J. M., & Yang, J.-B. (2017). A bibliometric analysis of Operations Research and Management Science. *Omega*, 73, 37-48. <https://doi.org/10.1016/j.omega.2016.12.004>
- [32] Bartol, T., Budimir, G., Dekleva-Smrekar, D., Pusnik, M., & Juznic, P. (2013). Assessment of Research Fields in scopus and web of science in the view of National Research Evaluation in Slovenia. *Scientometrics*, 98(2), 1491-1504. <https://doi.org/10.1007/s11192-013-1148-8>
- [33] AlRyalat, S. A., Malkawi, L. W., & Momani, S. M. (2019). Comparing bibliometric analysis using pubmed, Scopus, and web of science databases. *Journal of Visualized Experiments*, (152). <https://doi.org/10.3791/58494-v>
- [34] Barlow, J., & Venables, T. (2004). Will technological innovation create the true lifetime home? *Housing Studies*, 19(5), 795-810. <https://doi.org/10.1080/0267303042000249215>
- [35] Zhao, L., Tang, Z.-ying, & Zou, X. (2019). Mapping the knowledge domain of smart-city research: A Bibliometric and Scientometric analysis. *Sustainability*, 11(23), 6648. <https://doi.org/10.3390/su11236648>
- [36] Portet, F., Vacher, M., Golanski, C., Roux, C., & Meillon, B. (2013). Design and evaluation of a smart home voice interface for the elderly: Acceptability and objection aspects. *Personal and Ubiquitous Computing*, 17(1), 127-144. <https://doi.org/10.1007/s00779-011-0470-5>
- [37] Peek, S. T. M., Luijkx, K. G., Rijnaard, M. D., Nieboer, M. E., van der Voort, C. S., Aarts, S., van Hoof, J., Vrijhoef, H. J. M., & Wouters, E. J. M. (2016). Older adults' reasons for using technology while aging in place. *Gerontology*, 62(2), 226-237. <https://doi.org/10.1159/000430949>
- [38] Robinson, H., MacDonald, B., & Broadbent, E. (2014). The role of healthcare robots for Older People at Home: A Review. *International Journal of Social Robotics*, 6(4), 575-591. <https://doi.org/10.1007/s12369-014-0242-2>
- [39] Pal, D., Funilkul, S., Charoenkitkarn, N., & Kanthamanon, P. (2018). Internet-of-things and smart homes for elderly healthcare: An end user perspective. *IEEE Access*, 6, 10483-10496. <https://doi.org/10.1109/access.2018.2808472>
- [40] Bansal, P., & Kockelman, K. M. (2016). Are we ready to embrace connected and self-driving vehicles? A case study of Texans. *Transportation*, 45(2), 641-675. <https://doi.org/10.1007/s11116-016-9745-z>
- [41] Mital, M., Chang, V., Choudhary, P., Papa, A., & Pani, A. K. (2018). Adoption of internet of things in India: A test of competing models using a structured equation modeling approach. *Technological Forecasting and Social Change*, 136, 339-346. <https://doi.org/10.1016/j.techfore.2017.03.001>
- [42] Courtney, K., Demiris, G., Rantz, M., & Skubic, M. (2008). Needing Smart Home Technologies: The perspectives of older adults in continuing care retirement communities. *Journal of Innovation in Health Informatics*, 16(3), 195-201. <https://doi.org/10.14236/jhi.v16i3.694>
- [43] Demiris, G., Oliver, D. P., Dickey, G., Skubic, M., & Rantz, M. (2008). Findings from a participatory evaluation of a smart home application for older adults. *Technology and Health Care*, 16(2), 111-118. <https://doi.org/10.3233/thc-2008-16205>
- [44] Courtney, K. L. (2008). Privacy and senior willingness to adopt Smart Home Information Technology in Residential Care Facilities. *Methods of Information in Medicine*, 47(01), 76-81. <https://doi.org/10.3414/me9104>
- [45] Liu, J., Guo, X., Xu, S., Song, Y., & Ding, K. (2023). A new interpretation of scientific collaboration patterns from the perspective of symbiosis: An investigation for long-term collaboration in publications. *Journal of Informetrics*, 17(1), 101372. <https://doi.org/10.1016/j.joi.2022.101372>
- [46] Zhai, L., & Yan, X. (2022). A directed collaboration network for exploring the order of scientific collaboration. *Journal of Informetrics*, 16(4), 101345. <https://doi.org/10.1016/j.joi.2022.101345>
- [47] Pan, R. K., & Saramäki, J. (2012). The strength of strong ties in scientific collaboration networks. *EPL (Europhysics Letters)*, 97(1), 18007. <https://doi.org/10.1209/0295-5075/97/18007>
- [48] Mahmood, B., A. Sultan, N., H. Thanoon, K., & S. Kadhim, D. (2021). Measuring scientific collaboration in co-authorship networks. *IAES International Journal of Artificial Intelligence (IJ-AI)*, 10(4), 1103. <https://doi.org/10.11591/ijai.v10.i4.pp1103-1114>
- [49] Peek, S. T. M., Wouters, E. J. M., van Hoof, J., Luijkx, K. G., Boeije, H. R., & Vrijhoef, H. J. M. (2014). Factors influencing acceptance of technology for aging in place: a systematic review. *International journal of medical informatics*, 83(4), 235-248. <https://doi.org/10.1016/j.ijmedinf.2014.01.004>
- [50] Almusaylim, Z. A., & Zaman, N. (2018). A review on Smart Home Present State and challenges: Linked to context-awareness internet of things (IOT). *Wireless Networks*, 25(6), 3193-3204. <https://doi.org/10.1007/s11276-018-1712-5>
- [51] Layton, N., & Steel, E. (2019). The convergence and mainstreaming of Integrated Home Technologies for people with disability. *Societies*, 9(4), 69. <https://doi.org/10.3390/soc9040069>
- [52] Suh, S., Kim, B.-S., & Chung, J. H. (2015). Convergence Research Directions in cognitive sensor networks for elderly housing design. *International Journal of Distributed Sensor Networks*, 11(9), 196280. <https://doi.org/10.1155/2015/196280>
- [53] Rodriguez-Garcia, P., Li, Y., Lopez-Lopez, D., & Juan, A. A. (2023). Strategic decision making in smart home ecosystems: A review on the use of Artificial Intelligence and internet of things. *Internet of Things*, 22, 100772. <https://doi.org/10.1016/j.iot.2023.100772>
- [54] Yao, L., Li, J., & Li, J. (2020). Urban Innovation and Intercity Patent Collaboration: A network analysis of China's National Innovation System. *Technological Forecasting and Social Change*, 160, 120185. <https://doi.org/10.1016/j.techfore.2020.120185>
- [55] Ehrenhard, M., Kijl, B., & Nieuwenhuis, L. (2014). Market adoption barriers of multi-stakeholder technology: Smart Homes for the Aging Population. *Technological Forecasting and Social Change*, 89, 306-315. <https://doi.org/10.1016/j.techfore.2014.08.002>
- [56] Turjamaa, R., Pehkonen, A., & Kangasniemi, M. (2019). How smart homes are used to support older people: An integrative review. *International Journal of Older People Nursing*, 14(4). <https://doi.org/10.1111/opn.12260>
- [57] Philip, S. J., Luu, T. (Jack), & Carte, T. (2023). There's no place like home: Understanding users' intentions toward securing internet-of-things (IOT) smart home networks. *Computers in Human Behavior*, 139, 107551. <https://doi.org/10.1016/j.chb.2022.107551>
- [58] Nilashi, M., Abumalloh, R. A., Samad, S., Alrizq, M., Alyami, S., Abosaq, H., Alghamdi, A., & Akib, N. A. (2022). Factors impacting customer purchase intention of Smart Home Security Systems: Social Data Analysis Using Machine Learning Techniques. *Technology in Society*, 71, 102118. <https://doi.org/10.1016/j.techsoc.2022.102118>
- [59] Carnemolla, P. (2018). Ageing in place and the internet of things – how smart home technologies, the built environment and caregiving intersect. *Visualization in Engineering*, 6(1). <https://doi.org/10.1186/s40327-018-0066-5>
- [60] Hošnjak, A. M., & Pavlović, A. (2021). Older adults knowledge about using smart technology during the COVID-19 crisis-a qualitative pilot study. *IFAC-PapersOnLine*, 54(13), 675-679. <https://doi.org/10.1016/j.ifacol.2021.10.529>

# Automatic Generation of Image Caption Based on Semantic Relation using Deep Visual Attention Prediction

M. M. EL-GAYAR 

Department of Information Technology-Faculty of Computers and Information, Mansoura University  
Mansoura 35516, Egypt

Faculty of Computer Science and Engineering, New Mansoura University, New Mansoura, Egypt

**Abstract**—While modern systems for managing, retrieving, and analyzing images heavily rely on deriving semantic captions to categorize images, this task presents a considerable challenge due to the extensive capabilities required for manual processing, particularly with large images. Despite significant advancements in automatic image caption generation and human attention prediction through convolutional neural networks, there remains a need to enhance attention models in these networks through efficient multi-scale features utilization. Addressing this need, our study presents a novel image decoding model that integrates a wavelet-driven convolutional neural network with a dual-stage discrete wavelet transform, enabling the extraction of salient features within images. We utilize a wavelet-driven convolutional neural network as the encoder, coupled with a deep visual prediction model and Long Short-Term Memory as the decoder. The deep Visual Prediction Model calculates channel and location attention for visual attention features, with local features assessed by considering the spatial-contextual relationship among objects. Our primary contribution is to propose an encoder and decoder model to automatically create a semantic caption on the image based on the semantic contextual information and spatial features present in the image. Also, we improved the performance of this model, demonstrated through experiments conducted on three widely used datasets: Flickr8K, Flickr30K, and MSCOCO. The proposed approach outperformed current methods, achieving superior results in BLEU, METEOR, and GLEU scores. This research offers a significant advancement in image captioning and attention prediction models, presenting a promising direction for future work in this field.

**Keywords**—*Semantic image captioning; deep visual attention model; long short-term memory; wavelet driven convolutional neural network*

## I. INTRODUCTION

One of the active research topics in the field of computer vision is the creation of captions for images automatically. Image captioning refers to generate a text-based description or caption for a given image. This task unites computer vision and natural language processing methodologies to produce an easily understandable narrative that concisely conveys the image's content. The primary objective is to offer a brief and precise depiction of the elements, settings, actions, and occurrences depicted in the image [1-3]. There is an increasing daily demand for image retrieval and analysis systems because they are used in many fields and on a large scale via the

Internet, social media, and various search engines [4] [5]. Some ways in which image captioning benefits daily life include the following:

- For individuals with visual impairments: Image captions offer crucial details about an image's content, allowing them to gain a better understanding of the context surrounding the image.
- Cross-lingual communication: By translating image captions into various languages, individuals from different language backgrounds can better grasp the content of an image, promoting intercultural communication and appreciation.
- Search Engine Optimization (SEO): By offering pertinent textual data connected to an image, image captions can enhance SEO. This allows search engines to index and rank the content more precisely, improving online visibility and discoverability.
- User engagement on social media: Image captions contribute to a better user experience on social media platforms by supplying contextual information and additional details about images. This results in increased interaction and improved communication among users.
- Educational and e-learning contexts: Image captions play a supportive role in learning environments by making visual content more explicit and accessible for students, especially those facing learning disabilities or language challenges. This assistance leads to enhanced learning outcomes and better understanding of diverse topics.
- Data management and retrieval: Image captioning assists in organizing and locating visual information within extensive databases, simplifying the process of searching for particular images or content based on their descriptions.

In summary, image captioning serves as a crucial tool that enhances accessibility, comprehension, and distribution of visual content, providing advantages to a broad spectrum of users and applications in everyday life. This type of research is a vital topic that researchers are attracted to because it

combines three main areas: machine learning, natural language processing, and computer vision. It also serves a wide range of practical applications. Fusing these components result in sophisticated systems capable of autonomously interpreting the situation depicted in the image and generating coherent sentences to describe it.

The conventional method for image captioning consists of two key components: feature extraction and language modeling. In the feature extraction stage, an input image is processed by a pre-trained Convolutional Neural Network (CNN) model, such as VGG, Inception, or ResNet, to derive high-level visual features. Subsequently, during the language modeling phase, these extracted features are input into a language model, typically a Recurrent Neural Network (RNN) or a Long Short-Term Memory (LSTM) network, which produces a word sequence that forms the caption. By training the language model on a vast dataset of images and their associated captions, it learns to associate visual features with right words and phrases. Automatic image caption creation can be used in many practical systems and applications such as image retrieval through search engines, video labels, answering visual questions, assisting visually impaired people, biomedical imaging, robotics, etc. Recently, multiple approaches have been developed to automatically generate image captions, reducing many computer vision challenges [6].

The first method used in the image retrieval process relies on comparing the input image with a similar template to create a caption for the image through the matching or comparison process. However, the effectiveness of this method remains unproven, and it yields imprecise outcomes when dealing with intricate images containing multiple targets. Consequently, an alternative strategy relies on the development of a deep neural network, where the image is encoded, and captions are produced using a language model. In this process, the visual content of the image is analyzed in depth, and then this information is translated into natural language text descriptions. Nevertheless, there is a need to enhance CNN-based attention models by effectively utilizing multi-scale features in this model.

This paper introduces an automatic image captioning framework that generates semantically meaningful captions. The approach uses a deep neural network architecture, comprising a CNN that encodes the visual features and RNN that decodes and generates the text [7], [8]. It then employs LSTM [9], [10] and gated recurrent units (GRU) [11] to derive

significant insights. The key contributions of this article can be summarized as follows:

- Propose an encoder and decoder framework to automatically create a semantic caption on the image based on the semantic contextual information and spatial features present in the image.
- A Deep Visual Prediction Model (DVPM) is proposed by enabling the extraction of further semantic information from the image to utilize the convolutions on the feature maps generated using the Wavelet-driven Convolutional Neural Network (WCNN). Both channel and spatial attention are calculated using this approach, which are derived from the resulting feature maps.
- A semantic spatial contextual connection derived from the WCNN model is established to predict area proposals between distinct objects within the image.
- The feature maps produced by the WCNN model are leveraged by a Semantic Relation Extractor (SRE) to predict region proposals to determine the spatial relationships among diverse objects in the image.

The effectiveness of the suggested framework is assessed by employing three widely recognized benchmark datasets: Flickr8K, Flickr30K, and MS-COCO. Furthermore, a comparison with existing studies is conducted using various evaluation metrics, including Bilingual Evaluation Understudy (BLEU), Metric for Evaluation of Translation with Explicit Ordering (METEOR), and Consensus-based Image Description Evaluation (CIDEr).

This manuscript is organized into five sections. Section II reviews relevant previous work. Section III describes the proposed approach in detail. Experiments and results are presented in Section IV. Section V concludes the paper.

## II. RELATED WORKS

This section will review related studies regarding recent modalities in image captioning using attention mechanisms. A previously trained CNN (Encoder) would generate a hidden state (HS) in classic image captioning. Next, decoding this hidden state utilizes an LSTM (as the decoder) to frequently generate each word from the state. However, when the model attempts to produce the next word of the caption, there is an issue that this word typically only describes part of the image.

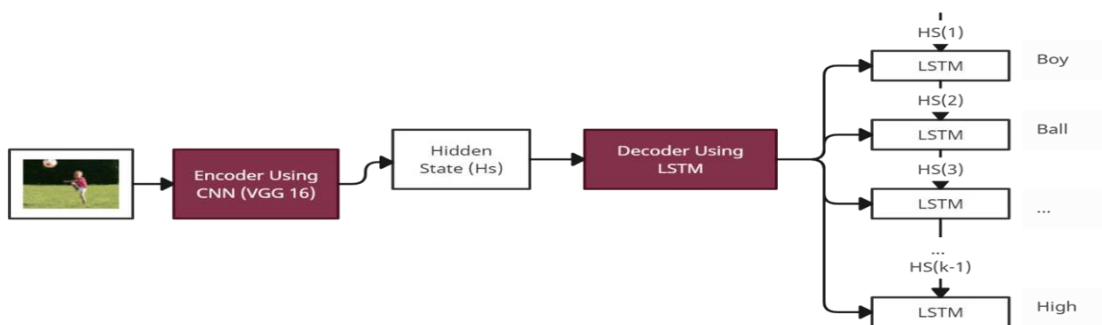


Fig. 1. A traditional image captioning model.



It is also unable to capture the essence of the entire input image. The model cannot efficiently generate different words for different parts of the image. Therefore, the attention mechanism is useful for representing the image [12]–[14]. Thus, generating an appropriate textual description requires a deeper understanding of the image's spatial and semantic content. As previously mentioned, initial efforts to create image translations involve extracting visible features using RAM (CRF) and converting these features into text through holistic or consensus-based improvements. Later recovery methods generate translations of single or multiple sentences from predefined phrases based on visual similarities. Developing a deep neural network architecture aids in achieving more advanced visual and natural language modeling by producing more insightful descriptions of the image.

Karpathy et al. [15] proposed a multimodal RNN for producing better descriptions using an alignment of replacement between the segments of the image and the sentence. Deng et al. [16] proposed an adaptive attention model with a visual sentinel. This model is presented to extract the global image characteristics of the encoding phase.

Zha et al. [17] proposed a context-aware visual policy network for better caption generation, reducing the dependency on previously predicted words using fine-grained image sentence captioning. Yu et al. [18] proposed a model that used dual attention (P and D) feature maps in the hierarchical image to explore the visual semantic connections and improve the quality of the sentences created.

Yang et al. [19] proposed a CaptionNet model for improved caption generation, which decreased the reliance on previously predicted words. This model only allows attended image features to be input into the memory of CaptionNet through input gates. Cornia et al. [20] proposed an innovative image captioning technique utilizing memory vectors and connecting the encoder and decoder sections of the transformer model.

Li X et al. [21] proposed an innovative technique for aligning the image and language modalities to gain more reasonable semantic extraction from images using anchor points. Jiang et al. [22] introduced a Multi-Gate Attention model to enhance caption generation, expanding upon the conventional self-attention mechanism by integrating an extra Attention Weight Gate. Wang et al. [30] proposes an automatic architecture search method for neural networks focused on cross-modality tasks like image captioning. The method approximates the associative connection between visual and language models through the internal structure of RNN cells. Over 100 generated RNN variants exceed performance of 100 on CIDEr and 31 on BLEU4, with the top model achieving 101.4 and 32.6, respectively. Wu et al. [31] introduced a novel global-local discriminative objective built on a reference model to generate more detailed descriptive captions. Evaluated on MS-COCO, the method outperforms baselines significantly and competes well with top approaches. Self-retrieval experiments demonstrate its ability to generate discriminative captions. Wang et al. [32] introduced a visual attention layer for low-level visual information and a semantic

attention layer for high-level semantic attributes. The margin-based loss encourages more discriminative captions. Extensive experiments on COCO and Flickr30K datasets validate the approach, demonstrating superior performance in captioning. The method achieves state-of-the-art 70.6 CIDEr-D on Flickr30K and competitive 123.5 CIDEr-D on COCO.

Although these methods effectively generate image captions, they fail to incorporate refined semantic components and the contextual spatial connections between various objects within the image. Therefore, expansions in network structure are essential to remedy these deficiencies. Furthermore, when several objects exist in the image, it is critical to properly consider the optical contextual connection between them to produce a more detailed and representative caption. This problem can be resolved by integrating attention mechanisms and assessing the spatial relations among elements within the instance. Table I shows the summary of recent related works.

TABLE I. SUMMARY OF RECENT PREVIOUS RELATED WORKS

Ref.	Dataset	BLEU	METEOR	CIDEr
7	FLICKER 8K	21.5	20.8	-
16	FLICKER 8K	25.7	22.6	52.6
19	FLICKER 8K	21.3	20.4	-
27	FLICKER 8K	16	-	-
16	FLICKER 30K	22.3	19.6	-
19	FLICKER 30K	19.8	18.5	-
30	FLICKER 30K	24.9	20.9	59.7
31	MS-COCO	37.2	28.4	123.4
32	MS-COCO	36.2	27.8	121.1

### III. PROPOSED FRAMEWORK

As illustrated in Fig. 1, a typical image captioning model would use a pre-trained convolutional neural network (CNN), such as VGG-16, to encode the input image and generate image features (HS) [23], [24]. Then, it would decode this HS using a Long Short-Term Memory (LSTM) and recursively render each caption word. The downside of this approach is that when the model tries to generate the following word in the caption, it fails to fully comprehend the overall meaning or essence of the entire input image. Therefore, a semantic deep visual attention mechanism can be helpful. With a semantic deep attention mechanism, the image is separated into  $n$  regions, and we calculate with CNN representations of each region  $HS(1), \dots, HS(n)$ . When the RNN-decoder generates a further word, the attention procedure concentrates on the appropriate region of the image, so the decoder only uses exact areas. Attention could be considerably distinguished into two types [25]–[29]:

- Global Attention is positioned on all origin positions, as shown in Fig. 2.
- Local Attention is positioned just on a few sources' places, as shown in Fig. 3.

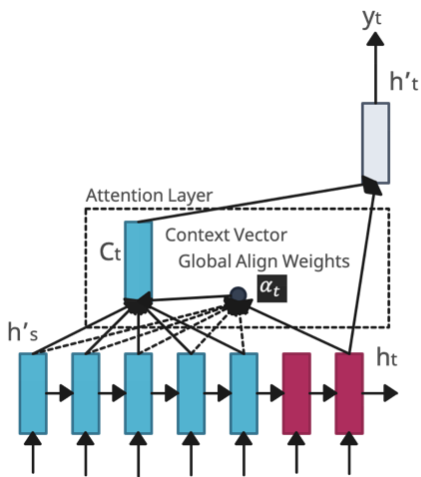


Fig. 2. Global attention model.

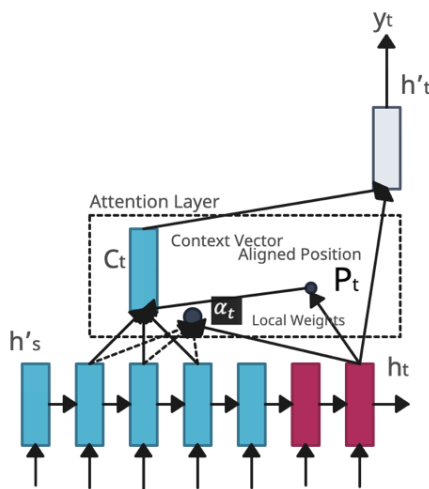


Fig. 3. Local attention model.

The global attention considers each HS coding for the excitation of the context vector. However, focusing global attention on all the main collateral terms of all destiny words is computationally expensive. In addition, it is not valid to use long phrases. To address this limitation, we can employ local attention to focus only on a small, relevant subset of the image features (HS) for generating each word in the caption. The proposed framework consists of an encoder- decoder model with a Deep Visual Prediction Model (DVPM), that transforms an input image (IMG) into a series of encoded expressions and words,  $T = [T_1, T_2, \dots, T_L]$ , with  $T_i \in \mathbb{R}^M$ , depicting the image, where  $L$  is the rendered caption's size, and  $M$  is the terminology size. The architecture of the proposed framework is illustrated in Fig. 6. The proposed framework is divided into four main phases. The first phase is the encoder using WCNN. The second phase is DVPM. The third phase is the semantic relation extractor. The final phase is the decoder using the LSTM model.

#### A. Encoder Phase

Comprising the WCNN model, the encoder merges two tiers of different wavelet decomposition alongside

convolutional neural network layers to extract the image's visual characteristics, as illustrated in Fig. 4. The Level 1 and Level 2 features obtained from the CNN layers are bilinearly downsampled and fused into a  $32 \times 32 \times 960$  feature map. In the first phase, the input image (I) is first resized to  $256 \times 256$  dimensions. The image is then separated into RGB color components. Each color component is decomposed into specifics and approximations using low-pass (LP) and high-pass (HP) discrete wavelet filters. The implementation of dual-phase discrete wavelet decomposition generates  $\{LP, LF\}$ ,  $\{HP, LF\}$ ,  $\{LP, HF\}$ , and  $\{HP, HF\}$  sub-bands, where LF and HF represent the low-frequency and high-frequency sections of the input image, respectively. In the second phase, only the  $\{LP, LF\}$  sub-band encounters further disassembly for each of the three elements. These components are combined and fused at every tier with the initial dual CNN stage outputs, encompassing four layers featuring numerous convolutional and pooling layers with a  $2 \times 2$  kernel dimension, as shown in Fig. 5.

Table II offers detailed information on the different convolutional layers. By incorporating the DWT stage alongside CNN, we aim to improve the visual modeling of the input image and extract some unique spectral characteristics. This method assists in capturing finer details of objects, including spatial orientation and color information, which allows for the identification of visually salient features or regions within the image. These features draw more attention, much like the human visual system, due to their distinct characteristics compared to other areas.

#### B. DVPM Phase

Extracting semantic attributes from input image feature maps, including aspects like an object's scale, shape, and texture features, is crucial. Differences in these characteristics within an image can create obstacles to accurate identification or recognition. To generate a semantic feature map of dimensions  $32 \times 32 \times 256$ , four multi-receptive filters are employed: one consisting of 64 filters with a  $1 \times 1$  kernel size and the other three featuring  $3 \times 3$  kernel sizes, each containing 64 filters with dilation rates of 3, 5, and 7, respectively. An example of attention changes to reflect the relevant parts of the image is shown in Fig. 7.

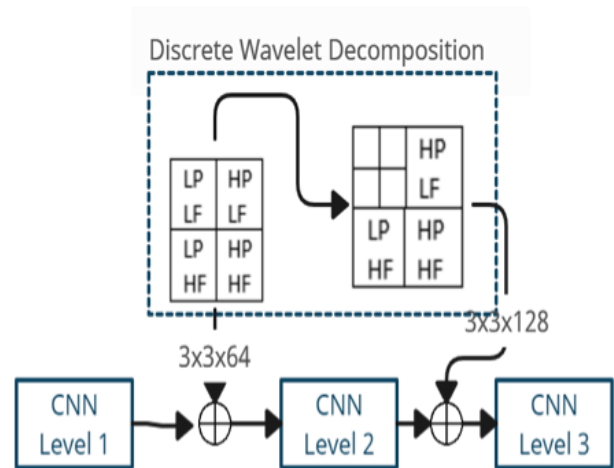


Fig. 4. Proposed encoder using WCNN.

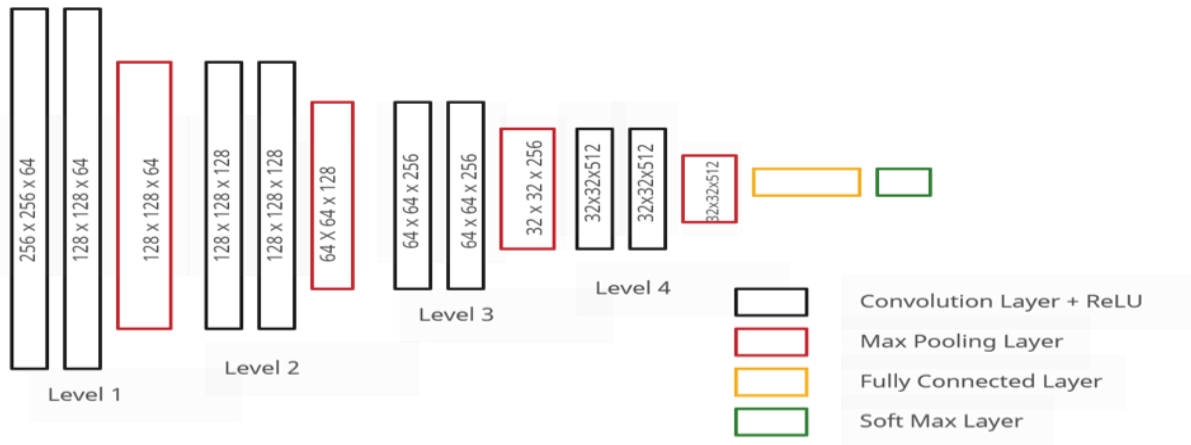


Fig. 5. Proposed CNN architecture.

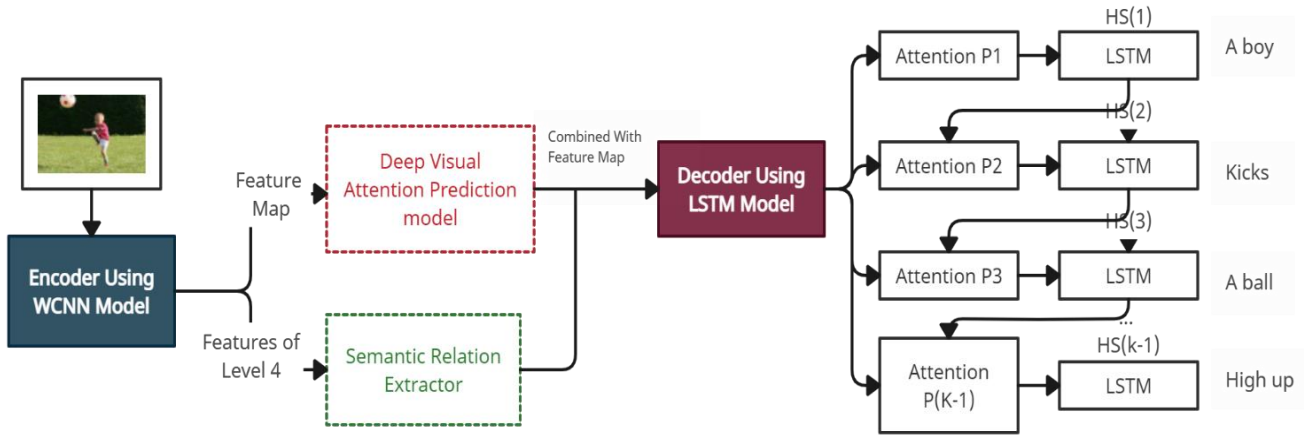


Fig. 6. Proposed framework.

TABLE II. SUMMARY OF RECENT PREVIOUS RELATED WORKS

Levels	Name	Kernel Size	Filter Size	Output Size
L1	Convolution L1,1 Convolution L1,2 Max Pool L1,1	3x3 3x3 2x2	64 64 64	256x256x64 256x256x64 128x128x64
L2	Convolution L2,1 Convolution L2,2 Max Pool L2,1	3x3 3x3 2x2	128 128 128	128x128x128 128x128x128 64x64x128
L3	Convolution L3,1 Convolution L3,2 Max Pool L3,1	5x5 5x5 2x2	256 256 256	64x64x256 64x64x256 32x32x256
L4	Convolution L4,1 Convolution L4,2	7x7 7x7	512 512	32x32x512 32x32x512

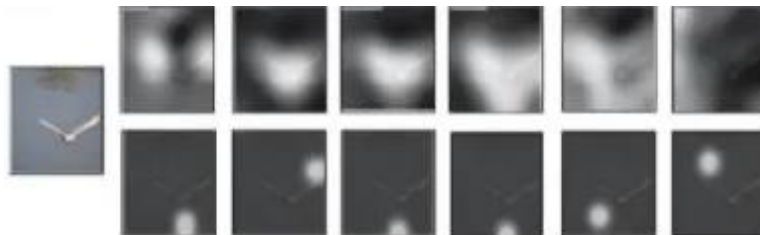


Fig. 7. Example of attention changes over time.

### C. Semantic Relation Extractor Phase

Generating rich image translations requires leveraging the contextual spatial relationships between multiple objects in the image and their semantic details. In the WCNN model, the ultimate tier's feature map serves as input for the Semantic Relation Extractor (SRE) to identify object regions in the image. Subsequently, objects are paired, and numerous uniformly sized 32x32 sub-images are created through resizing. Each sub-image is fed into the CNN layers that contain 64 filters, with each filter having a receptive field of 3x3. This process generates features that represent the spatial relationships between pairs of objects. The feature maps of individual objects are combined to form a 32x32x64-dimensional contextual spatial relation feature map. This feature map is then merged with the output from a feed-forward neural network of local attention (fa) and supplied to the LSTM to produce the next word in the caption.

### D. Decoder Phase

To generate more precise image captions, the channel attention weights  $W_c$  and spatial attention weights  $W_s$  are computed based on  $H_{t-1} \in R^n$ .  $D$  is the dimension of the hidden state. By using this method, additional contextual data is integrated into the image while generating captions. The feature map, denoted as  $F_{map}$  has dimensions  $F_{map} \in R^{h \times w \times c}$  where  $h$ ,  $w$ , and  $c$  represent the height, width, and a total number of channels of the feature map, respectively. The initial step involves average pooling on a per-channel basis, resulting in a channel feature vector,  $F_v \in R^c$ .

Since global attention focuses on all the words of the secondary origin of all objective words, This process becomes expensive and impractical for translating lengthy sentences. So instead, local attention concentrates on a small subset of the encoder's hidden states for each target word to address this limitation. So, we do softmax to get the input probability distribution of the channel attention weights ( $W_c$ ).  $W_c$  can be calculated as follows:

$$W_c = \text{Softmax}(E_{it}) \quad (1)$$

$$E_{it} = X_c(P_{t-1}, H_i) \quad (2)$$

$$X_c = V_{att}^T * \tanh(U_{att} * H_i + W'_c * P_t) \quad (3)$$

Where:

- $E_{it}$  means at every  $t^{th}$  time steps of decoder, how important  $i^{th}$  is the pixel location in the input image.
- $P_{t-1}$  is the pervious state of decoder.
- $H_i$  is the state of encoder.
- $X_c$  is simple feed forward neural network which is a linear transformation of input ( $U_{att} * H_i + W'_c * P_t$ ) and then a non-linearity (tanh) on the top of that.
- $V_{att}^T \in R^D$ ,  $U_{att} \in R^{K \times D}$ ,  $W'_c \in R^{K \times D}$ ,  $H_i \in R^D$  and  $P_t \in R^D$ .

Now, we need to feed weighted sum combination to decoder. So, the weighted sum of input (context vector  $C_t$ ) is calculated from Eq. (4).

$$C_t = \sum_{i=1}^T W_c H_i \text{ such that } \sum_{i=1}^T W_c = 1 \quad (4)$$

### E. Summary

In this part, we summarize the steps of the proposed system and link them to the proposed algorithms.

- 1) *Clean* data (as discussed in algorithm-I), i.e., clearing punctuations and numeric values from the text.
- 2) *Preprocessing* the images and captions (as discussed in algorithm-II and algorithm-III, respectively, by appending '<start>' and '<end>' labels to every caption) so that the proposed model understands the starting and stopping of each caption.
- 3) *We* have to reshape every image before feeding it to the WCNN model.
- 4) *The* captions will be tokenized, and a vocabulary of words in our data corpus will be established.
- 5) *Producing* Encoder Hidden States — The encoder employs a WCNN model that integrates dual-level discrete wavelet decomposition with CNN layers, efficiently extracting an image's visual features.
- 6) *Applying* DVMP to output the semantic feature map of size 32x32x256, we use four multi-receptive filters.
- 7) *Applying* SRE to find the object regions in the image by entering the feature map from the last level of the WCNN model.
- 8) *As* described in Algorithm 4, the RNN Local Decoder utilizes the hidden state (HS) from the previous decoder and the current decoder output.

The Decoder RNN processes these inputs to generate a new hidden state.

- 9) *The* alignment scores are calculated as in algorithm IV.
- 10) *Softmaxing* the previous scores.
- 11) *A* context vector is calculated.
- 12) *The* context vector is merged with the decoder's hidden state (HS), produced in Step 8, resulting in a new output.
- 13) *Steps* 6 through 13 are iteratively executed for each time step in the decoder until a token is generated.

<b>Algorithm I - Data Cleaning</b>
Input: Original Text (OT) Output: Cleaned Text (CT)
Start Procedure
OT ← Original Text
CT ← null
CT ← OT.translate(string.punctuation)
TL ← txt_length_more_than_1
TL ← null
Foreach word in CT.split():
IF length(word) > 1:
TL += " " + word
End IF
End Foreach
End Procedure

**Algorithm II - Image Preprocessing**

Input: Data  
Output: IMG

Start Procedure

```
[ ] ← IMG_vector
Foreach fnames in data["filename"]:
    path ← img_dir + "/" + fnames
    all_img_name_vector.append(path)
    IMG ← tf.io.read(path)
    IMG ← tf.image.decode_jpg(IMG, ch=3)
    IMG ← tf.image.resize(IMG, (224,224))
End Foreach
End Procedure
```

**Algorithm III - Caption Preprocessing**

Input: Data  
Output: Total Captions (t\_cp)

Start Procedure

```
[ ] ← t_cp
Foreach cp in data["cp"] astype(str):
    cp ← '>start<' + cp + '>end<'
    t_cp.append(cp)
End Foreach
End Procedure
```

**Algorithm IV - RNN Local Decoder**

Input: units, vocab\_size, features map (features) and hidden  
Output: state, attention weights (att\_w), Context Vector (CV)

Start Procedure

```
Uatten ← tf.keras.layers.Dense(units)
Wc ← tf.keras.layers.Dense(units)
Vatten ← tf.keras.layers.Dense(1)
hidden_time_axis ← tf.expand_dim(hidden, 1)
score ← use equation 3
att_w ← tf.softmax(score,axis=1)
att_w ← use equation 1
CV ← attention_weights * feature
CV ← use equation 4
End Procedure
```

#### IV. EXPERIMENTAL RESULTS ANALYSIS AND DISCUSSION

This section will discuss and compare the outcomes of our diverse experiments with pertinent prior studies. We implemented the proposed framework using TensorFlow 2.3 and executed it on Google Cloud with the help of Google Colab.

##### A. Description of Datasets

Numerous open-source datasets, including Flickr 8k, Flickr 30k, and MS COCO, are accessible for this research. The experiments are carried out on the following three benchmark datasets:

- Flickr 8k — 6400 images (training set), 700 images (validation set), and 700 images (testing set).
- Flickr 30k— 24k images (training set), 3k images (validation set), and 3k images (testing set).

- MS-COCO — 128k images (training set), 16k images (validation set), and 16k images (testing set).

##### B. Hyperparameters

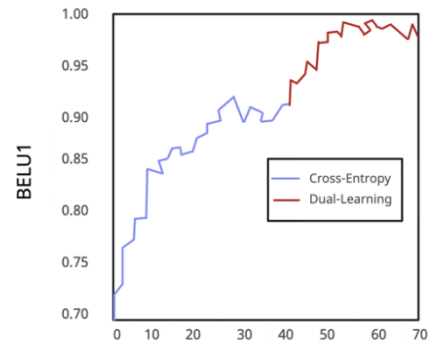
During the model's training stage, we used hyperparameter settings such as batch size, dropout, etc. The values of a few hyperparameters include the exponential decay rates for ADAM optimizer, learning rate, batch size, and dropout. The number of iterations used is 50. These hyperparameters are changed on a trial-and-error. Finally, the hyperparameters are tuned into our method to improve the results.

- For Flickr8k, Flickr30k, and MSCOCO datasets, *the batch sizes* employed are 16, 32, and 64, respectively.
- *Dropout*: To prevent overfitting, a dropout rate of 0.2 is applied, L2 regularization, and a weight decay value of 0.001.
- *Epochs*: The model begins with 40 epochs of training based on cross-entropy loss. Afterward, an extra 80 epochs of fine-tuning are conducted via dual learning to reach the highest CD score within the validation set.

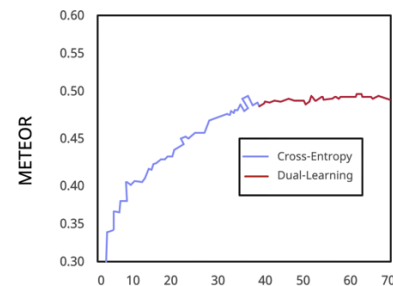
##### C. Evaluation Metrics

Our experiments use the performance evaluation metrics – BLEU as B score from 1 to 4, METEOR as MR score, and CIDEr as CD score. The BLEU metric is employed to assess the generated captions for the test set. Recognized as a reliable metric, BLEU quantifies the similarity between a single predicted sentence and multiple reference sentences. Table III provides a summary of the metrics featured in this paper. Additionally, the Beam search technique was utilized to evaluate the captions.

##### D. Results



(a) BLEU-1



(b) METEOR

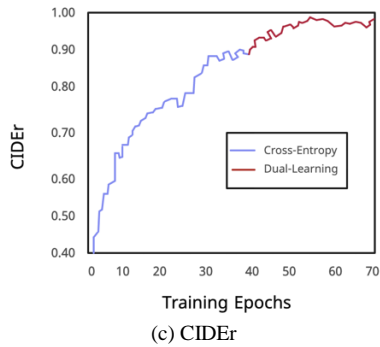


Fig. 8. Learning curves of the proposed framework according to each metric.



**Ground Truth:**  
Lone man climbing high on snowy mountain  
**Beam Search:**  
man is standing on top of snow mountain  
**BLEU Score**  
67.9



**Ground Truth:**  
Brown dog with tongue out and it is running through the grass  
**Beam Search:**  
Brown dog is running through a grassy field  
**BLEU Score**  
73.9



**Ground Truth:**  
Children playing in public waterspouts  
**Beam Search:**  
Children playing with water  
**BLEU Score**  
70.8

Fig. 9. Samples of RNN-generated captions on test images.

### E. Comparison Between the Proposed Model and the Previous Related Works

This sub-section displays the evaluation outcomes for the comparative analysis between the proposed model and the related previous works on different datasets.

TABLE III. COMPARISON BETWEEN THE PROPOSED MODEL AND RELATED WORKS ON THE FLICKER 8K DATASET

Ref.	B-1	B-2	B-3	B-4	MR	CD
7	64.7	45.9	31.7	21.5	20.8	-
27	57.9	38.3	24.5	16	-	-
19	67.2	45.9	31.4	21.3	20.4	-
16	68.1	49.3	34.9	25.7	22.6	52.6
Our Model	73.4	52.3	36.9	29.2	27.3	68.4

TABLE IV. COMPARISON BETWEEN THE PROPOSED MODEL AND RELATED WORKS ON THE FLICKER 30K DATASET

Ref.	B-1	B-2	B-3	B-4	MR	CD
7	64.6	44.8	30.7	20.5	17.8	-
27	57.3	36.9	24.1	15.7	15.3	-
19	66.9	43.9	29.6	19.8	18.5	-
16	66.2	46.7	32.5	22.3	19.6	-
30	-	-	-	24.9	20.9	59.7
Our Model	72.2	50.3	35.7	27.4	21.9	66.8

TABLE V. COMPARISON BETWEEN THE PROPOSED MODEL AND RELATED WORKS ON THE MS-COCO DATASET

Ref.	B-1	B-2	B-3	B-4	MR	CD
7	67	49.2	35.7	26.3	22.6	80.3
27	62.7	45.3	32.3	23.4	20.2	66.2
19	71.9	50.8	35.8	25.1	23.1	-
31	-	-	-	37.2	28.4	123.4
32	78.9	62.9	48.9	36.2	27.8	121.1
Our Model	79.8	63.4	50.1	39.2	28.8	123.9

### F. Discussion

As shown in Fig. 8 and Fig. 9, after about 40 epochs, all the evaluation metrics converge, and the performance of the proposed model evolves better when we fine-tune the model on the unpaired data by employing the dual learning mechanism. The results comparing the proposed model on the Flickr8K and Flickr30K datasets are presented in Table III and Table IV. As seen in Table III, the proposed model shows notable improvements of 2.3%, 2.8%, and 2.1% in B-1, B-4, and MR scores for the Flickr8K dataset. Likewise, the model achieves increases of 2.4% and 0.9% in B-4 and MR scores for the Flickr30K dataset in Table IV. The model also attains a respectable CD value of approximately 66.8. Table V displays the evaluation outcomes for the comparative analysis on the MSCOCO test partition. As indicated in Table IV, the proposed model yields a strong CD score of 123.9 and exhibits relative enhancements of around 0.9%, 0.5%, and 0.7% in B-4, MR, and CD scores, respectively. We can use the proposed model in IoT systems in [33],[34] to ensure controllability, safety and effectiveness as a future work. Unlike other methods, this improvement stems from the proposed model's image feature maps incorporating spectral information alongside spatial and semantic details. The model

can obtain detailed data during object identification by integrating discrete wavelet decomposition into the CNN model. Additionally, the model considers the contextual spatial relationships between objects in the image and employs spatial and channel-specific attention to enhance feature maps resulting from convolution. Using multi-receptive field filters facilitates the detection of visually prominent objects with diverse shapes, scales, and sizes.

## V. CONCLUSION

This manuscript introduces a deep visual attention framework for image caption generation, utilizing an encoder-decoder architecture based on semantic relationships. The encoder comprises a WCNN, while the decoder comprises a DVPM and LSTM. The DVPM calculates channel and location attention for visual features, taking into account the spatial-contextual relationship between various objects. Merging wavelet decomposition with the convolutional neural network allows the model to extract spatial, semantic, and spectral data from the input images. In-depth image captions are produced by applying spatial and channel-wise attention to the feature maps generated by DVPM and considering the contextual spatial relationships among objects via the CSE network. Assessments are conducted on three standard datasets—Flickr8K, Flickr16K, and MS-COCO—utilizing evaluation metrics such as BLEU, METEOR, and CIDEr. With the MS-COCO dataset, the model achieves remarkable B-4, MR, and CD scores of 39.2, 28.8, and 123.9, respectively. We believe there are several promising directions for future work. First, the proposed model could be refined and tested with various other attention mechanisms, potentially improving the model's performance even further. Second, the application of this model to other vision-and-language tasks, such as visual question answering and image-based storytelling, IoT systems could be explored. Additionally, the integration of other types of contextual information, such as object-object interaction or more explicit spatial information may enhance the model's ability to generate even more detailed and accurate captions. Finally, while the model has been tested on standard datasets, it would be worthwhile to evaluate its performance on a diverse array of real-world images and scenarios. These future research directions will help to further reinforce and extend the significant contributions of our study.

## REFERENCES

- [1] M. al Sulaimi, I. Ahmad, and M. Jeragh, "Deep Image Captioning Survey: A Resource Availability Perspective," Conference of Open Innovation Association, FRUCT, vol. 2021-May, pp. 3–13, May 2021.
- [2] H. Sharma, M. Agrahari, S. K. Singh, M. Firoj, and R. K. Mishra, "Image Captioning: A Comprehensive Survey," 2020 International Conference on Power Electronics and IoT Applications in Renewable Energy and its Control, PARC 2020, pp. 325–328, Feb. 2020.
- [3] S. Sukhi, A. Q. Ohi, M. S. Rahman, and M. F. Mridha, "A Survey on Bengali Image Captioning: Architectures, Challenges, and Directions," 2021 International Conference on Science and Contemporary Technologies, ICSCT 2021, 2021.
- [4] M. El-Gayar, H. Soliman and N. Meky, "A comparative study of image low level feature extraction algorithms", Egyptian Informat. J., vol. 14, no. 2, pp. 175-181, 2013.
- [5] M. M. El-Gayar, N. E. Mekky, A. Atwan and H. Soliman, "Enhanced search engine using proposed framework and ranking algorithm based on semantic relations", IEEE Access, vol. 7, pp. 139337-139349, 2019.
- [6] M. Stefanini, M. Cornia, L. Baraldi, S. Cascianelli, G. Fiameni, and R. Cucchiara, "From Show to Tell: A Survey on Deep Learning-based Image Captioning," IEEE Trans Pattern Anal Mach Intell, Jan. 2022.
- [7] A. Hani, N. Tagougui, and M. Kherallah, "Image caption generation using a deep architecture," Proceedings - 2019 International Arab Conference on Information Technology, ACIT 2019, pp. 246– 251, Dec. 2019.
- [8] C. S. Kanimozhiselvi, V. Karthika, S. P. Kalaivani, and S. Krithika, "Image Captioning Using Deep Learning," 2022 International Conference on Computer Communication and Informatics, ICCCI 2022, 2022.
- [9] X. Jia, E. Gavves, B. Fernando, and T. Tuytelaars, "Guiding Long-Short Term Memory for Image Caption Generation," Sep. 2015.
- [10] S. Wang et al., "Cascade attention fusion for fine-grained image captioning based on multi-layer LSTM," ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, vol. 2021-June, pp. 2245–2249, 2021.
- [11] L. Gao, X. Wang, J. Song, and Y. Liu, "Fused GRU with semantic-temporal attention for video captioning," Neurocomputing, vol. 395, pp. 222– 228, Jun. 2020.
- [12] P. Shah, V. Bakrola, and S. Pati, "Image captioning using deep neural architectures," Proceedings of 2017 International Conference on Innovations in Information, Embedded and Communication Systems, ICIIECS 2017, vol. 2018-January, pp. 1–4, Jan. 2018.
- [13] A. Hani, N. Tagougui, and M. Kherallah, "Image caption generation using a deep architecture," Proceedings - 2019 International Arab Conference on Information Technology, ACIT 2019.
- [14] I. Hrga and M. Ivašić-Kos, "Deep image captioning: An overview," 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2019 - Proceedings, pp. 995–1000, May 2019.
- [15] A. Karpathy and L. Fei-Fei, "Deep Visual-Semantic Alignments for Generating Image Descriptions," IEEE Trans Pattern Anal Mach Intell, vol. 39, no. 4, pp. 664–676, Dec. 2014.
- [16] M. Yang et al., "Multitask learning for cross-domain image captioning," IEEE Trans Multimedia, vol. 21, no. 4, pp. 1047–1061, Apr. 2019.
- [17] Z. J. Zha, D. Liu, H. Zhang, Y. Zhang, and F. Wu, "Context-Aware Visual Policy Network for Fine-Grained Image Captioning," IEEE Trans Pattern Anal Mach Intell, vol. 44, no. 2, pp. 710–722, Jun. 2019.
- [18] L. Yu, J. Zhang, and Q. Wu, "Dual Attention on Pyramid Feature Maps for Image Captioning," IEEE Trans Multimedia, vol. 24, pp. 1775–1786, 2022.
- [19] L. Yang, H. Wang, P. Tang, and Q. Li, "CaptionNet: A Tailor-made Recurrent Neural Network for Generating Image Descriptions," IEEE Trans Multimedia, vol. 23, pp. 835–845, 2021.
- [20] M. Cornia, M. Stefanini, L. Baraldi, and R. Cucchiara, "Meshed-Memory Transformer for Image Captioning," Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 10575–10584, Dec. 2019.
- [21] X. Li et al., "Oscar: Object-Semantics Aligned Pre-training for Vision-Language Tasks," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 12375 LNCS, pp. 121–137, Apr. 2020.
- [22] W. Jiang, X. Li, H. Hu, Q. Lu, and B. Liu, "Multi-Gate Attention Network for Image Captioning," IEEE Access, vol. 9, pp. 69700–69709, 2021.
- [23] S. J. Rennie, E. Marcheret, Y. Mroueh, J. Ross, and V. Goel, "Self-critical Sequence Training for Image Captioning," Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, vol. 2017-January, pp. 1179–1195, Dec. 2016.
- [24] H. Wang, H. Wang, and K. Xu, "Evolutionary recurrent neural network for image captioning," Neurocomputing, vol. 401, pp. 249–256, Aug. 2020.

- [25] K. Xu et al., "Show, Attend and Tell: Neural Image Caption Generation with Visual Attention," 32nd International Conference on Machine Learning, ICML 2015, vol. 3, pp. 2048–2057, Feb. 2015.
- [26] J. Wu, T. Chen, H. Wu, Z. Yang, G. Luo, and L. Lin, "Fine-Grained Image Captioning with Global- Local Discriminative Objective," *IEEE Trans Multimedia*, vol. 23, pp. 2413–2427, Jul. 2020.
- [27] M. A. Al-Malla, A. Jafar, and N. Ghneim, "Image captioning model using attention and object features to mimic human image understanding," *J Big Data*, vol. 9, no. 1, pp. 1–16, Dec. 2022.
- [28] P. Anderson et al., "Bottom-Up and Top-Down Attention for Image Captioning and Visual Question Answering," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 6077–6086, Jul. 2017.
- [29] Z. Deng, Z. Jiang, R. Lan, W. Huang, and X. Luo, "Image captioning using DenseNet network and adaptive attention," *Signal Process Image Commun*, vol. 85, p. 115836, Jul. 2020.
- [30] H. Wang, H. Wang, and K. Xu, "Evolutionary recurrent neural network for image captioning," *Neurocomputing*, 401:249–56, 2020.
- [31] J. Wu, T. Chen, H. Wu, Z. Yang, G. Luo, and L. Lin, "Fine-grained image captioning with global-local discriminative objective," *IEEE Trans Multimedia*, 23:2413–27, 2021.
- [32] S. Wang, Y. Meng, Y. Gu, L. Zhang, X. Ye, J. Tian, L. Jiao, "Cascade attention fusion for fine-grained image captioning based on multi-layer lstm," 2021 IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 2245–2249, 2021.
- [33] H. Fetooh, M.M.El-Gayar, A. Aboelfetouh, "Detect Technique and Mitigation Against a Phishing Attack". *Int J Adv Comput International Journal of Information ManagementSci Appl. The Science and Information (SAI) Organization*, 12:177–88, 2021.
- [34] N.A. Hikal, M.M. El-Gayar, "Enhancing IoT botnets attack detection using machine learning-IDS and ensemble data preprocessing technique", *Lect Notes Networks Syst, Springer*, 114:89–102, 2020.



# Enhancing Oil Price Forecasting Through an Intelligent Hybridized Approach

Hicham BOUSSATTA<sup>1</sup>, Marouane CHIHAB<sup>2</sup>, Younes CHIHAB<sup>3</sup>, Mohammed CHINY<sup>4</sup>  
Laboratory of Computer Sciences, Faculty of Sciences Ibn Tofail University, Kenitra, Morocco<sup>1,2,3</sup>  
Faculty of Sciences Ibn Tofail University, Kenitra, Morocco<sup>4</sup>

**Abstract**—The oil market has long experienced price fluctuations driven by diverse factors. These shifts in crude oil prices wield substantial influence over the costs of various goods and services. Moreover, the price per barrel is intricately intertwined with global economic activities, themselves influenced by the trajectory of oil prices. Analyzing oil behavior stands as a pivotal means for tracking the evolution of barrel prices and predicting future oil costs. This analytical approach significantly contributes to the field of crude oil price forecasting. Researchers and scientists alike prioritize accurate crude oil price forecasting. Yet, such endeavors are often challenged by the intricate nature of oil price behavior. Recent times have witnessed the effective employment of various approaches, including Hybrid and Machine Learning techniques to address similarly complex tasks, though they often yield elevated error rates, as observed in financial markets. In this study, the goal is to enhance the predictive precision of several weak supervised learning predictors by harnessing hybridization, particularly within the context of the crude oil market's multifaceted variations. The focus extends to a vast dataset encompassing CPSE Stock ETF prices over a period of 23 years. Ten distinct models, namely SVM, XGBoost, Random Forest, KNN, Gradient Boosting, Decision Tree, Ridge, Lasso, Elastic Net, and Neural Network, were employed to derive elemental predictions. These predictions were subsequently amalgamated via Linear Regression, yielding heightened performance. The investigation underscores the efficacy of hybridization as a strategy. Ultimately, the proposed approach's performance is juxtaposed against its individual weak predictors, with experiment results validating the findings.

**Keywords**—Oil market; prediction; crude oil; hybrid approach; CPSE stock ETF price; machine learning; stock markets

## I. INTRODUCTION

Hybridization approach refers to the combination of two different approaches or models to improve the accuracy of exchange rate forecasting in time series analysis. [1], given the significance of time-series prediction in many real-world situations, it is important to carefully select an appropriate model. For this reason, numerous performance measures have been proposed in the literature [1-7] to assess forecast accuracy and compare different models. These are known as performance metrics [6]. The goal of time series models is to gain an understanding of the underlying factors and structure that shaped the observed data, fit a model, and uses it for forecasting. These models have a wide range of applications in the daily operations of electric utilities, such as energy generation planning, energy purchasing, load switching, and contract evaluation [8], The purpose of forecasting is to make

and improve decisions, increase profits, and in the case of forecasting oil prices, better decisions largely depend on the accurate prediction of trends, actual prices, and expected prices  $x(t)$  and  $x'(t)$ . The ability to predict movement can be measured statistically (R2) [9]. Similarly, the significance of forecasting lies in reducing the risk or uncertainty involved in short-term decision making and planning for long-term growth. Forecasting the demand and sales of a company's products usually begins with a macroeconomic forecast of the overall level of economic activity, such as Gross National Product (GNP). Companies use macroeconomic forecasts of general economic activity as inputs for their microeconomic estimates of the demand and sales for the industry and the firm. The demand and sales for a business are typically estimated based on its historical market share and planned marketing strategy (e.g, forecasting by product line and region). Companies use long-term forecasts for the industry and the economy to determine the necessary investment in plant and equipment to achieve their long-term growth objective. The focus of this study is on multi-step ahead prediction of crude oil prices. This involves extrapolating the crude oil price series by predicting multiple time-steps into the future without access to future outputs. Despite the influence of many complex factors, oil prices exhibit highly non-linear behavior, making it challenging to predict future oil prices, especially when looking several steps ahead (Fan et al., 2008) [10]. The unpredictability of crude oil prices is due to their sensitivity to fluctuations in both global demand and supply. The world economy was destabilized for a decade when oil supplies were disrupted 40 years ago. The formation of the OPEC cartel and the nationalization of the oil industries in the Middle East led to a quadrupling of world oil prices and caused steep recessions in the mid-1970s. The 1979 overthrow of the Shah of Iran by Muslim clerics disrupted Iran's oil supplies, leading to another round of even deeper recessions. The productivity of the future oil market and the expected accuracy of future prices are evaluated. The precision of forecasts using futures prices is compared to that of other methods, including time series and econometric models, as well as key forecasts. The predictive power of futures prices is further investigated by comparing the forecasting accuracy of end-of-month prices with weekly and monthly averages, using different weighting systems [11] Previous studies have shown that the behavior of oil prices is non-linear and traditional econometric and statistical methods struggle to provide accurate predictions in these cases. To address this issue, newer techniques like genetic algorithms, artificial neural networks, and support vector machines have emerged [13]. Alizadeh and Mafinezhad [13] used a General

Regression Neural Network (GRNN) model to forecast Brent oil prices by incorporating seven types of variables as inputs. The authors claimed that this model performed well under various conditions and provided a high level of accuracy. Previous studies have shown that oil price behavior is non-linear, and traditional econometric and statistical models may not be sufficient for analyzing this behavior [12]. To address this, new techniques such as genetic algorithms, artificial neural networks, and support vector machine have emerged [13]. Alizadeh and Mafinezhad [13] used a General Regression Neural Network (GRNN) model to forecast Brent oil price, incorporating seven types of structures as inputs. They found that their model provided a high level of accuracy under challenging conditions. Predicting oil prices is a challenging task due to its significant impact on various economic and non-economic aspects. There is currently a lack of consensus among experts on the most effective methods and models for forecasting oil prices. To address this issue, a hybridization approach that combines multiple models can be used to increase forecasting accuracy [14]. In this study, the aim is to enhance the accuracy of crude oil price predictions by combining weak predictors through hybridization. The dataset comprises daily CPSE Stock ETF prices spanning 23 years. Ten different machine learning models are utilized. (SVM, Random Forest, Gradient Boosting, Neural Network, XGBoost, Decision Tree, Ridge, Lasso, Elastic Net and KNN) to make individual predictions, and then combine these predictions using a Linear Regression model to achieve enhanced performance. The results clearly illustrate the advantages of employing the hybridization approach. After testing the 10 models, The SMRM approach yielded the most accurate results, as it converges towards the minimum of the empirical response and minimizes information loss, outperforming the other models. The study's findings suggest that the Hybrid Proposed System (SMRM) stands out as the most efficient option, outperforming all other individual models. The SMRM achieved an average negative MAPE of -0.023, which was the highest among all models and sets. The proposed SMRM hybrid system offers a promising solution for predicting crude oil prices, leveraging the power of machine learning, and combining multiple models to better capture the complex relationships between different factors. The experiments reveal that SMRM excels over existing models in both accuracy and stability, making it a valuable tool for investors, traders, and other stakeholders in the energy sector. The system can also be continually refined and improved by incorporating irregular factors like political risks and extreme weather events, which can help to better predict changes in crude oil prices. With further development, this approach could have important implications for supporting decision-making and risk management in the energy sector, enabling stakeholders to make more informed and effective decisions in the dynamic and complex world of stock market trading. By providing more accurate and reliable predictions of crude oil prices, the SMRM hybrid system has the potential to revolutionize how we approach predicting crude oil prices, providing valuable insights that can help optimize decision-making and drive greater value in the energy sector. This passage highlights that the main research contribution is the development of the SMRM hybrid system, which combines

machine learning models to improve the accuracy and stability of crude oil price predictions. It also emphasizes the system's potential to enhance decision-making and risk management in the energy sector.

The rest of the paper is organized as follows: Section II reviews the current literature on forecasting crude oil prices. Following that, Section III details methodology, and Section IV covers the proposed approach. The empirical results are presented in section V.

## II. RELATED WORK

Crude oil prices are difficult to predict due to a complex pricing system with insufficient information, numerous variables, and inaccurate elements [15]. Despite this challenge, researchers are actively exploring methods to accurately forecast crude oil prices and manage related risks. While traditional methods like Arima and Arma have been used, they often fall short in the face of complex data and asymmetric effects. The growth of AI and text mining technologies provide new opportunities to predict crude oil prices and measure investment risk. One such successful machine learning model [16] uses artificial intelligence to predict crude oil prices, including the use of Decision Trees (DTs), a commonly used technique in crude oil modeling. To further improve accuracy, [17] incorporates technical trading indicators like RSI and Stochastic Oscillator into the Random Forest model for minimizing investment risk. These advances in AI technology have the potential to significantly improve the accuracy of crude oil price forecasts. Two PCA-KNN models, which combined PCA for information reduction and KNN for oil price forecasting, were tested on historical EUR/USD exchange rate data sets over a 10-year period. These models achieved the highest success rate of 77.58%. To improve the success rate, [18] presents a novel approach to financial time series prediction by combining K-Nearest Neighbor (KNN) Regression with Principal Component Analysis (PCA). The authors aim to enhance the accuracy of financial time series forecasting, a critical task in the finance domain, [19] proposed a PANK model, which involves three components: (1) Principal Component Analysis to minimize redundant information, (2) Affinity Propagation Clustering for feature extraction through example generation and corresponding cluster formation, and (3) nearest k-neighbour regression reformulated through nested regression for prediction modeling. The PANK model was tested on a 15-year historical data set of the Chinese stock market index, yielding a success rate of 80%. Previous studies have defined the behavior of oil prices as a statistical system, but these methods only provide logical results for linear behavior. The study in [20] used the SVM model to estimate oil prices, but these methods are inadequate for highly complex and non-linear data. The study in [21] recently compared different forecasting models, including ARIMA, FNN, ARFIMA, MS-ARFIMA, and the RW model, and found that the SVM model performed best and is a strong candidate for crude oil price forecasting in one or more stages. Machine learning models often require hand-crafted features, which can make them challenging to implement in real-world situations, especially with large amounts of data. A recent and successful approach comes from the subfield of machine learning, deep learning [22]. The

accuracy of forecasting can be improved through hybridization by combining simple forecasts from multiple weaker predictors [23]. A hybrid system combining Artificial Neural Networks (ANN) and Recurrent Neural Networks (RES) based on text mining was proposed to improve prediction performance [24]. Another study proposed a Multi-Intelligent Bat-Neural Network Multi-Agent System (BNNMAS) for predicting the price of oil-linked stocks, comparing it to genetic algorithm neural network (GANN) and generalized neural network regression (GRNN) [24]. However, both systems are subject to performance problems due to the identification of tuning parameters. A recent study [25] developed an intelligent system for forecasting oil prices using time series models, but the system is not suitable for predicting long-term trends. In reference to the study of oil prices, various methods have been used to make predictions and analyze the factors affecting its fluctuation. The research in [26] utilized the Complementary Empirical Ensemble Mode Decomposition (CEEMD) to break down the barrel price into its components and identify the impact of extreme events on crude oil prices. The researcher combined the Iterative Cumulative Sum of Squares (ICSS) test and Chow's test to detect structural breaks, then used ARIMA and SVM models to forecast oil prices. The results showed that the SVM-CEEMD-ARIMA model with structural breaks was the best performing model compared to SVM and ARIMA models alone. During the COVID-19 pandemic, [27] attempted to predict oil price movements using ANN and SVM models. The results showed that his model outperformed other ANN and SVM models, with an RMSE value of 0.6018 and a MAE value of 0.5295. The study in [28] used a Convolutional Neural Network (CNN) to extract features from online news texts, divided into categories such as oil price, oil production, overall oil consumption, and oil stocks. Other models such as MLR, BPNM, SVM, RNN, and LSTM were used for prediction and the results showed that social media factors play a role in oil price prediction. During the Russian-Ukrainian conflict, [29] forecasted oil prices by using exotic options such as Asian Options, Barrier Options, and Gap Options. The GARCH model and Monte Carlo simulation were used to study the options and the results showed that considering the overall performance of all exotic options was better. In [30], the authors analyzed the relationship between oil prices and various wars, including the first and second Gulf Wars and the Russian-Ukrainian War. The results showed that the relationship between real GDP growth and oil prices differed between periods and that it was possible to predict oil prices during the Russian-Ukrainian War. The ongoing conflict between Russia and Ukraine continues to have a significant impact on the financial market and oil prices. The research in [31] used the TVP-VAR technique to identify the sources of oil market volatility and the interconnections between gold, crude oil, and the stock market on February 24, 2022. The results showed that the conflict between Ukraine and Russia affects the interdependence of the markets analysed, both in stable and war situations. In [32], the authors compared the performance of support vector machines (SVM), K-nearest neighbors (KNN), and random forest (RF) models in predicting crude oil prices. The authors found that the SVM model outperformed the KNN and RF models in terms of both in-sample and out-of-sample prediction accuracy. Guliyev and Mustafayev in [33] to

compare their predictive accuracy and identify the most effective model for crude oil price forecasting. The three machine learning models used in the study are Linear Regression, Support Vector Regression (SVR), and Random Forest Regression. These models are trained using a range of explanatory variables, such as supply and demand factors, macroeconomic indicators, geopolitical risks, and oil market-specific variables, such as oil inventories. The study found that all three models can effectively predict the WTI crude oil price changes with reasonable accuracy. However, the Random Forest model produced the most accurate forecasts compared to the other two models. In addition, the study found that geopolitical risks, such as tensions between OPEC members and potential supply disruptions, have the most significant impact on WTI crude oil prices. The study's results suggest that machine learning models can be a useful tool for crude oil price forecasting. The findings could be useful for traders, investors, and policymakers who need to make informed decisions based on the expected future price dynamics of crude oil. The study in [34] is to evaluate the effectiveness of different machine learning models in predicting the closing prices of stocks. The study employs three machine learning algorithms: Random Forest (RF), Support Vector Regression (SVR), and Multilayer Perceptron (MLP) for predicting the closing prices of stocks using a range of input features such as volume, moving averages, and technical indicators. The study finds that all three machine learning models are effective in predicting the closing prices of stocks, with Random Forest performing the best, followed by Support Vector Regression and Multilayer Perceptron. The study also found that technical indicators, such as Relative Strength Index (RSI) and Moving Average Convergence Divergence (MACD), were the most effective input features for the prediction models. Overall, the study suggests that machine learning models can be useful for predicting the closing prices of stocks and can help traders and investors make informed decisions based on expected future prices. [35] is to develop a hybrid artificial intelligence model to predict the uniaxial compressive strength of oil palm shell concrete. The study uses a combination of machine learning algorithms, including Artificial Neural Networks (ANN), Genetic Programming (GP), and Support Vector Regression (SVR), to develop a hybrid model for predicting the uniaxial compressive strength of oil palm shell concrete. The model uses input features such as the water-binder ratio, curing time, and oil palm shell content. The study finds that the hybrid model outperforms individual machine learning models in predicting the uniaxial compressive strength of oil palm shell concrete. The hybrid model also achieves a high prediction accuracy with a coefficient of determination (R-squared) value of 0.965. The results of the study suggest that the hybrid model can be a useful tool for predicting the uniaxial compressive strength of oil palm shell concrete, which is important for the design and construction of sustainable building materials. The hybrid model can also be extended to predict the properties of other types of concrete by modifying the input features. The research in [36] is to propose a novel hybrid model for forecasting crude oil prices based on time series decomposition. The study combines two machine learning algorithms, Support Vector Regression (SVR) and Artificial Neural Networks (ANN), with a time series decomposition

method called Seasonal and Trend decomposition using Loess (STL), to create a hybrid model for crude oil price forecasting. The study finds that the proposed hybrid model outperforms individual machine learning models and traditional time series models in forecasting crude oil prices. The hybrid model achieves a high prediction accuracy with a Mean Absolute Percentage Error (MAPE) value of 3.22%. The results of the study suggest that the proposed hybrid model can be a useful tool for predicting crude oil prices, which is important for making informed decisions in the oil and gas industry. The study also highlights the importance of combining different machine learning algorithms and time series decomposition methods for improving the accuracy of crude oil price forecasting. The study in [37] is to propose a novel approach for predicting crude oil prices by combining complex network analysis and deep learning algorithms. The study uses a complex network analysis to identify the relationships and dependencies between different economic variables, such as exchange rates and stock prices, and crude oil prices. The identified network is then used as input for deep learning algorithms, specifically a Convolutional Neural Network (CNN) and a Recurrent Neural Network (RNN), to predict crude oil prices. The study finds that the proposed approach outperforms traditional machine learning models and time series models in predicting crude oil prices. The CNN and RNN models achieve high prediction accuracy with a Mean Absolute Error (MAE) value of 0.55 and 0.51, respectively. The results of the study suggest that the proposed approach can be a useful tool for predicting crude oil prices, which is important for making informed decisions in the oil and gas industry. The study also highlights the importance of considering the complex relationships and dependencies between different economic variables in crude oil price prediction. In [38], Abdollahi and Ebrahimi propose a new hybrid model for predicting the Brent crude oil price. The proposed model combines two different techniques: an ensemble of Extreme Learning Machines (ELMs) and a wavelet transform. The study first applies a wavelet transform to decompose the time series data into different frequency bands, which allows the model to capture different patterns and trends in the data. The decomposed signals are then used as input for the ELM ensemble, which is a machine learning technique that combines multiple ELM models to improve prediction accuracy. The study finds that the proposed hybrid model outperforms several benchmark models, including traditional statistical models, machine learning models, and other hybrid models. The proposed model achieves a high prediction accuracy, with a Mean Absolute Percentage Error (MAPE) value of 1.76% for one-day ahead forecasting and 3.31% for five-day ahead forecasting. The results of the study suggest that the proposed hybrid model can be an effective tool for predicting the Brent crude oil price, which is important for making informed decisions in the oil and gas industry. The study also highlights the importance of combining different techniques to improve prediction accuracy and to capture different patterns and trends in the data. The research in [39] is to propose a weighted hybrid data-driven model for forecasting daily natural gas prices. The proposed model combines two different techniques: an Empirical Mode Decomposition (EMD) method and a Support Vector Machine (SVM) method.

The study first applies the EMD method to decompose the original time series data into several Intrinsic Mode Functions (IMFs). These IMFs capture the different temporal scales of the natural gas prices and are used as input variables for the SVM method. The SVM method is a popular machine learning algorithm that can be used for regression and classification tasks. To further improve the performance of the model, the study introduces a weight-based approach that assigns different weights to the historical data based on their importance. The weights are calculated using a genetic algorithm that searches for the optimal weight values. The study finds that the proposed weighted hybrid data-driven model outperforms several benchmark models, including traditional statistical models, machine learning models, and other hybrid models. The proposed model achieves a high prediction accuracy, with a Mean Absolute Percentage Error (MAPE) value of 1.87% for one-day ahead forecasting and 2.55% for five-day ahead forecasting. The results of the study suggest that the proposed weighted hybrid data-driven model can be an effective tool for forecasting daily natural gas prices, which is important for making informed decisions in the energy industry. The study also highlights the importance of combining different techniques and introducing a weight-based approach to improve prediction accuracy.

Prediction accuracy, with a Mean Absolute Percentage Error (MAPE) value of 1.87% is for one-day ahead forecasting and 2.55% is for five-day ahead forecasting. The results of the study suggest that the proposed weighted hybrid data-driven model can be an effective tool for forecasting daily natural gas prices, which is important for making informed decisions in the energy industry. The study also highlights the importance of combining different techniques to improve prediction accuracy.

In this article, the aim is to enhance prediction accuracy by combining several weak predictors through hybridization to address the varying degrees of variability in the oil market. Ten models are used (SVM, XGBoost, Random Forest, KNN, Gradient Boosting, Decision Tree, Ridge, Lasso, Elastic Net and Neural Network) to obtain basic predictions and then integrate them in a Linear Regression for a better outcome. The previous section discussed different types of algorithms including Text Mining algorithms, genetic algorithms, and deep learning algorithms. Although advancements have been made in dynamic system modeling and analysis over the past 23 years to minimize prediction error, the uncertainty of learning models remains limited. However, combining diverse predictive models can prove to be effective in improving prediction accuracy. The study centers on the integration of various regression methods (SVM, Random Forest, Gradient Boosting, Neural Network, XGBoost, Decision Tree, Ridge, Lasso, Elastic Net and KNN) to make predictions and produce a final prediction through stacking. The idea is to construct a predictive model by combining various models, as shown in the diagram:

#### A. Units

The original training dataset, (X), consists of m observations and n features, resulting in an (m×n) matrix. Multiple models M are trained on X using a training method, such as cross-validation. These models make predictions for the result, (y), which are then consolidated into a second

training dataset,  $(X^{(2)})$ , with a shape of  $(m \times M)$ . These  $(M)$  predictions serve as features for this second-level dataset. The goal of creating a second-level model is to produce the final prediction by utilizing the combination of these different models.

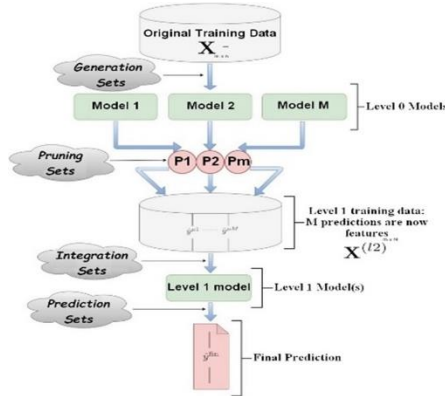


Fig. 1. Stacking Multi-Regression Models (SMRM).

### B. Pruning Sets

Evaluation criteria such as MAPE, MAE, RMSE and R-Squared  $R^2$  play a crucial role in selecting the best models based on ranking. These criteria are presented in Table I.

TABLE I. META-FEATURES USED

Features	Description
MAPE	Mean Absolute Percentage Error
MAE	Mean Absolute Error
RMSE	Root Mean Square Error
$R^2$	R-Squared

### C. Integration Sets

This section explains the behavior of the selected models as shown in Table II in their combinations for predicting barrel price. Stacking models is a technique used to create a secondary dataset for cross-validation using k-fold to address two key issues:

To make off-sample predictions, the stacking process uses predictions  $f_1, \dots, f_m$  to determine the generator biases for the learning set in different regions, where each model is most effective. The right linear combination with the weights vector is then learned by the meta-learner  $a_1, \dots, (i=1, \dots, m)$ :

$$f_{stacking}^{(x)} = \sum_{i=1}^m a_i f_i(x) \quad (1)$$

### D. Prediction Sets

The final predictions are generated from the training data  $X$  or from the second-level learner's model(s). The stacking model is used to select various sub-learners and to study how to collect and combine sub-models and their predictions. A meta-model is employed to merge the best predictions from all models. Each model provides predictions for the outcome  $(y)$ , which are integrated into a second training dataset  $(X^{(2)})$ , resulting in  $(m \times M)$  predictions. These second-level data possess  $M$  new characteristics. A second-level model is created to generate the results used for the final prediction. Fig. 1

illustrates the overall design of the results after applying this approach. Three models were generated and the basic model type at level 0, as well as the differences between the ten models, is explained as follows:

- 1) The first model used in the base layer is Random Forest
- 2) The second model used in the base layer is KNN
- 3) The third model used in the base layer is SVM
- 4) The fourth model used in the base layer is Gradient Boosting
- 5) The fifth model used in the base layer is Decision Tree
- 6) The sixth model used in the base layer is Ridge.
- 7) The seventh model used in the base layer is Lasso
- 8) The eighth model used in the base layer is Elastic Net
- 9) The ninth model used in the base layer is XGBoost
- 10) The tenth model used in the base layer is Neural Network

TABLE II. DATA MINING ALGORITHMS

Algorithm	Description
KNN	K-Nearest Neighbour
Decision Tree	Decision Tree
SVM	RSuport Vector Machine
Neural Network	Neural Network
Ridge	Ridge
Lasso	Lasso
Elastic Net	Elastic Net
XGBoost	XGBoost
Random Forest	Random Forest
Gradient Boosting	Gradient Boosting

## III. STUDY OF THE PROPOSED APPROACH

The prediction capacity is tested against some reference models. First, the data description will be presented in Section (A). Second, all measures for evaluating prediction performance and the statistical tests that compare and adjust predictive accuracy will be discussed in Section (B). Finally, the stacking learning sets algorithm will be explained in Section (C).

### A. Dataset

The ETF Prices data was used as a reference point and was uploaded to the ETF prices website [40]. This data represents the global oil price and is daily in nature, covering the period from 2000 to 2023 with 5751 observations. The data consists of important factors that impact supply and demand and the dependent variable of oil consumption. These variables were selected to model the barrel price series for the following reasons: Firstly, they are closely linked to oil prices and represent various drivers of the end price. Secondly, the relationship between these factors and the oil price series is noisy, non-linear, and volatile, but one of them is likely to provide valuable information on oil price schedules at any given time. Thirdly, more insights can be gained by including as many variables as possible. Finally, the system that contains

all these models, namely (SVM, Random Forest, Gradient Boosting, XGBoost, Neural Network, Decision Tree, Ridge, Lasso, Elastic Net and KNN) is mainly powerful in modeling high-dimensional data using all these variables. The data is divided into two parts, with the training samples consisting of the first 80% of observations of all series and the rest used as test data, as shown in Fig. 2. All data is obtained from websites, including the Energy Information Administration (EIA), Exchange Traded Funds (ETF), and Yahoo Finance (36). The visualization of the entire actual time series (annual and monthly) is shown in Fig. 3 and Fig. 4. For model formation, oil prices and exogenous variables are pre-processed using first differences helps remove zero values, and the use of standardized variables avoids estimation problems such as an explosion of parameters. The methods are designed to model price series rather than oil performance series. The proposed system SMRM is developed with the aim of identifying and validating the factors that contribute to oil price variations. To achieve this objective, a hybrid system SMRM is implemented and utilized to understand the fluctuation of the barrel price while utilizing information from Yahoo Finance. Table IV demonstrates the storage and testing of the data based on its quantitative characteristics. The system reveals 12 key performance indicators which were used as explanatory variables in the model with the expected next-day oil price as the dependent and output variable. These key performance indicators are listed in Table III.

TABLE III. KEY PERFORMANCE INDICATORS AFFECTING CRUDE OIL PRICES

Algorithm	Description
S_3	Moving average for past 3 days
S_9	Moving average for past 9 days
RSI_3	Moving average of Relative Strength Index for past 3 days
RSI_9	Moving average of Relative Strength Index for past 9 days
MME_26	Exponential Moving Average for past 26 days
MME_12	Exponential Moving Average for past 12 days
%K	Stochastic Oscillator
RVI	Relative Vigo Index
MOM	Momentum
MACD	Mobile Average Convergence Divergence
%D_3	Moving Average of %L for past 3 days
%D_9	Moving Average of %L for past 9 days

TABLE IV. OIL MARKET EXPLANATORY VALUES

Features	Description
Close	Reference to the end of a trading
Open	Reference to the starting period of trading
Hight	Reference to the involving large amounts of price
Low	Reference to the reaching of the price

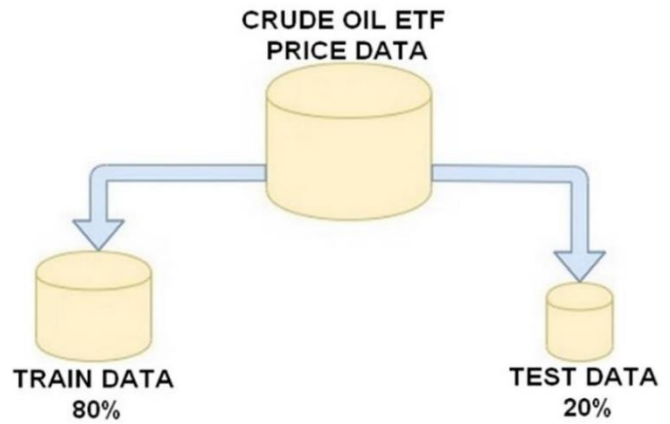


Fig. 2. Splitting the dataset.

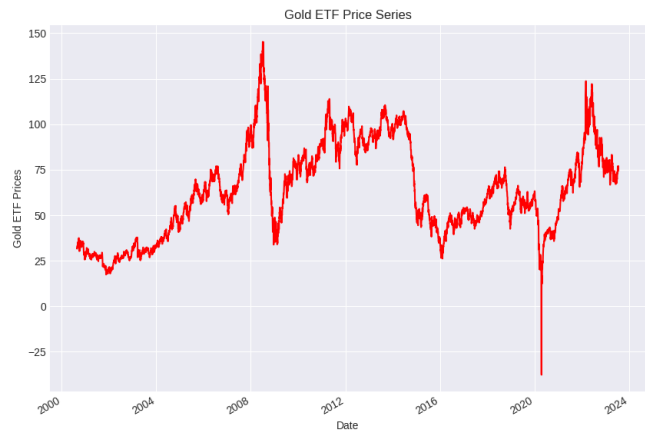


Fig. 3. Annual crude oil price.

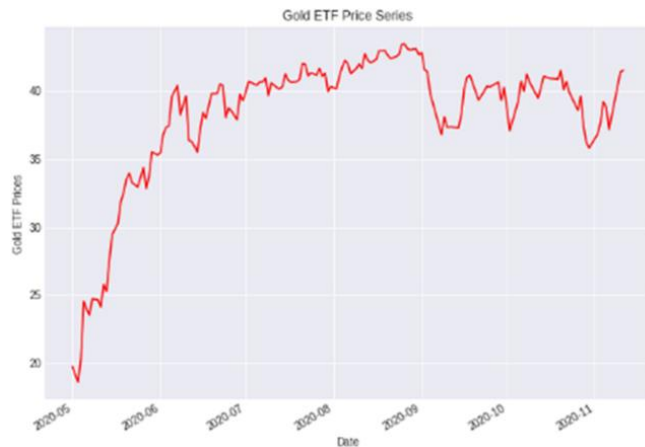


Fig. 4. Monthly crude oil price.

**B. Performance evaluation criteria and statistical test**

In evaluating the performance of the prediction models, four important indicators were used. These include the Mean Absolute Percentage Error MAPE calculated using Eq. (2), Mean Absolute Error MAE calculated using Eq. (3), Root Mean Squared Error (RMSE) calculated using Eq. (4) and R-Squared R2 calculated using Eq. (5). These indicators play a

crucial role in estimating the performance of prediction models across various aspects.

$$MAPE = \frac{1}{N} \sum_{t=1}^N \frac{y(t) - \hat{y}(t)}{\hat{y}(t)} \quad (2)$$

$$MAE = \frac{1}{N} \sum_{t=1}^N (y - \hat{y})^2 \quad (3)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (y(t) - \hat{y}(t))^2} \quad (4)$$

$$R2 = 1 - \frac{\sum_{t=1}^N (y - \hat{y})^2}{\sum_{t=1}^N (y - \bar{y})^2} \quad (5)$$

Or  $y(t)$  and  $\hat{y}(t)$  represent, respectively, the actual value and the predicted value,  $a(t)=1$  if  $(y(t+1) - y(t))(\hat{y}(t+1) - y(t)) \geq 0$  or  $a(t)=0$ , and  $N$  is the size of the predictions.

### C. The Algorithm for Staking Learning Sets

This section outlines the process of the proposed system, which is based on a typical sequence of the dataset for improved prediction. The general design of the method is expressed using pseudo code.

Input: Dataset  $\{D=(x_1,y_1),(x_2,y_2),\dots,(x_m,y_m)\}$

First-level learning algorithms  $L_1,L_2,\dots,L_n$

Second-level learning algorithm  $L$

Process:

% Have a training of first-level individual learner  $h_t$  by applying the first-level learning algorithm  $L_t$  to the original dataset  $D$

for  $t = 1, \dots, T$ :

$h_t=L_t(D)$

end

% Generation of a new dataset

$D^{\wedge}=\emptyset$

for  $t = 1, \dots, m$

for  $t = 1, \dots, T$

$z_{it}=h_{it}(x_i)$  % Used  $h_t$  to predict training

dataset  $x_i$

end

$D^{\wedge}=D^{\wedge} \cup \{(z_{i1}, z_{i2}, \dots, z_{iT}), y_i\}$

end

% Have a training of the second  $h^{\wedge}$  learner by applying the second level, learning the  $L$  algorithm to the new dataset set  $D^{\wedge}$

## IV. RESULTS

### A. Benchmarking and Comparison of the Predictive Modelling

The results of testing 10 models (SVM, XGBoost, Random Forest, KNN, Gradient Boosting, Decision Tree, Ridge, Lasso,

Elastic Net and Neural Network) show that the use of the Random Forest algorithm is more effective and closer to reality. This is because the Random Forest algorithm minimizes information loss and converges towards the minimum of empirical response, as shown in Fig. 6. On the other hand, the other models appear to be less effective and less modulable, with more variability present in the data, as seen in Fig. 5 and 7 in which, the case of the two SVM and KNN models is considered.

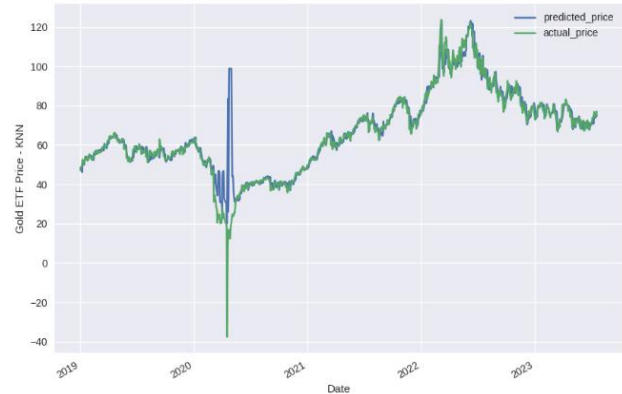


Fig. 5. Predicted and actual ETF Price via the KNN model.



Fig. 6. Predicted and actual ETF price via the Random Forest regressor model.

### B. Examination of Algorithms

First, the ETF Price data was utilized as a reference dataset, and various machine learning models were examined on the dataset. The assessment will include the following 10 algorithms.

- 1) KNN
- 2) Random Forest
- 3) SVM
- 4) XGBoost
- 5) Gradient Boosting
- 6) Neural Network
- 7) Decision Tree
- 8) Ridge
- 9) Lasso
- 10) Elastic Net

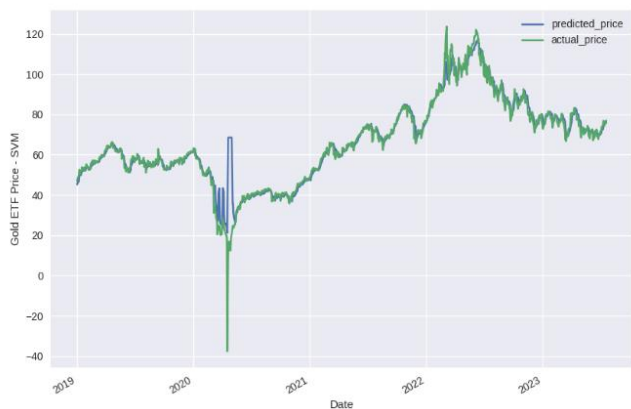


Fig. 7. Predicted and actual ETF Price via the SVM model.

All the algorithms will be evaluated, and their average performance will be compared by describing the distribution of accuracy scores for each. The models will undergo evaluation based on the MAPE. Due to the stochastic nature of the algorithm and numerical accuracy differences, the results may vary. In this context, it was found that the Random Forest algorithm gave the best result with a negative MAPE of approximately -0.021, as shown in Table VI, and with a best R2 of 95.40% as shown in Fig. 8.

### C. The Combination of Models

The approach defines the Stacking Multi-Regression Models (SMRM) by initially presenting a list of tuples for the 10 basic models and subsequently defining the Linear Regression, which acts as a meta-model, to combine the predictions of the basic models and learn how to best combine the outputs of each of the 10 separate models (see Fig. 1). This implementation allows us to assess the performance of each model and the findings indicate that SMRM is the most efficient when compared to the other models, with a negative MAPE of approximately -0.023. The average and median scores for the SMRM are the highest in comparison to the other individual models, as seen in Table V. However, A stacking set can be chosen as the final model, fine-tune it, and use it to make predictions on novel datasets using the linear model created from all the training data. The prediction method estimates the ETF Price (y) for the explanatory variable X as shown in Fig. 10. The results show that the R2 has a score of 97% as shown in Fig. 9. A score close to 100% indicates that the model effectively explains the ETF prices for crude oil. The cumulative returns, represented as a purchase signal, are shown in Fig. 11, where a "1" value indicates that the expected price for the next day is higher than the current day's price, while no position is taken otherwise.

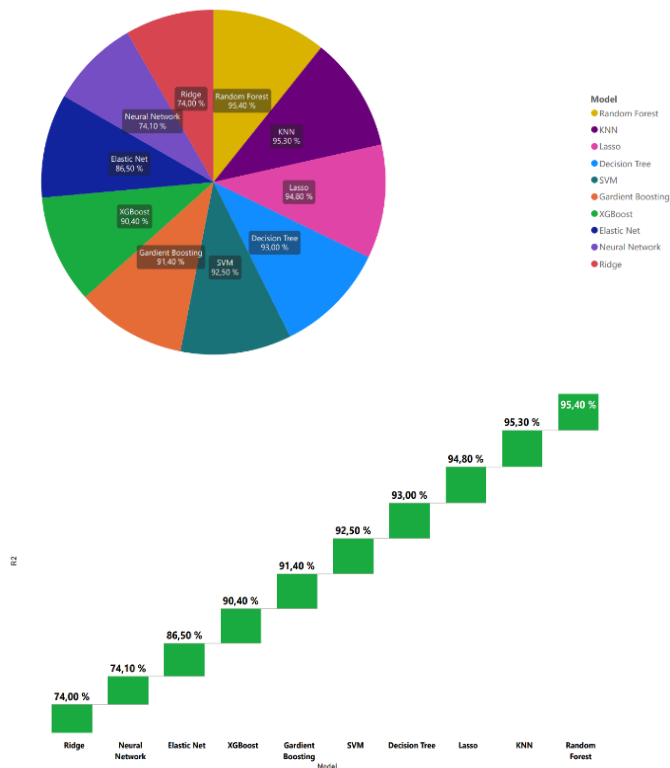


Fig. 8. Average performance of algorithms.

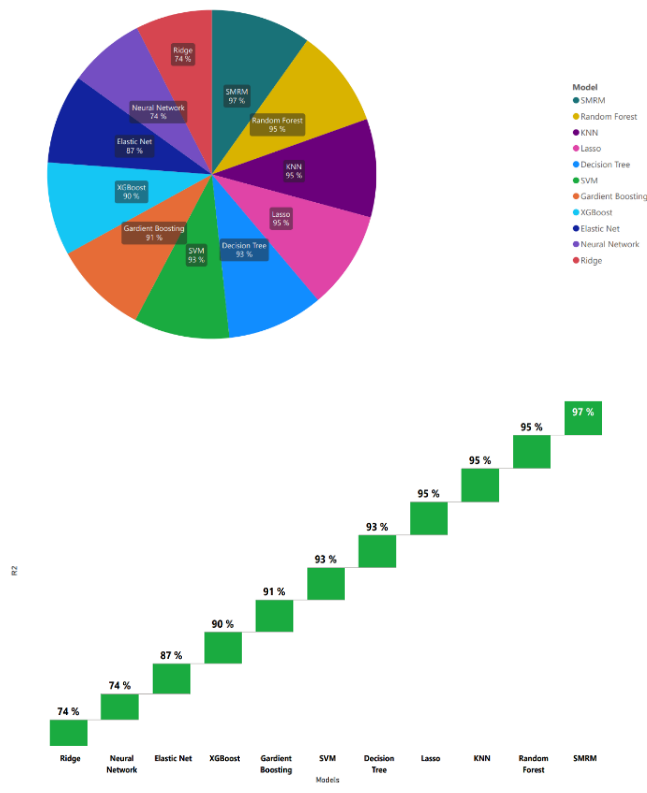


Fig. 9. Average algorithm performance and SMRM.





Fig. 10. Predicted price and ETF price.

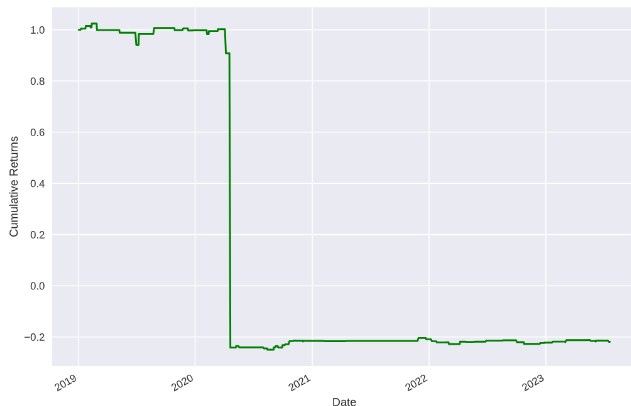


Fig. 11. Cumulative returns signals crude oil price.

## V. DISCUSSION

In this section, a stacking set can be chosen as the final model. Both individual and overall predictors have been applied to address the problem of estimating the expected price of the next day, providing a more comprehensive evaluation of the stability of the SMRM hybrid system. All models were examined in their predictions. It is evident that the proposed SMRM hybrid system was the best for forecasting oil prices, as shown in Fig 10, compared to the other models studied. In all models, the Random Forest Regressor model based on the SMRM hybrid system not only achieved the highest accuracy in estimation, as measured by MAPE, but also achieved the highest success rate in R2, as measured by R2. Additionally, among the models studied, the KNN performed the worst in growth forecasts. This model not only had the lowest MAPE, but also achieved the worst score in terms of R-Squared, as measured by R2. This could be since SVM was a linear regressor, which couldn't capture non-linear models. In addition to the Random Forest, KNN, and the other models based on the SMRM hybrid system, which produced the best and worst results, respectively, all models studied produced encouraging mixed results, which were analyzed using four evaluation criteria (i.e., MAPE, MAE, RMSE, and R2). First, in terms of level accuracy, the results of the MAPE measurement showed that the Random Forest based on SMRM achieved the best results, followed by the other models that are based on SMRM was the weakest, as shown in Table VI. Second, high

accuracy does not necessarily imply a high success rate in predicting the R2. It is crucial that the R2 is correct for the policy maker to make an investment plan in oil-related processes (production, price, and demand). Thus, comparison of the R2 is essential. Similar conclusions can be drawn from Table VII regarding the R2. The SMRM hybrid system achieved significantly higher results and closer to 100 than all others, followed by the other 10 overall models (i.e., SVM, XGBoost, Random Forest, KNN, Gradient Boosting, Decision Tree, Ridge, Lasso, Elastic Net and Neural Network). The 10 overall methods typically outperformed individual forecasting models, and among the overall methods, the Random Forest model based on the SMRM hybrid system produced the best results, while the other models based on SMRM, The Ridge model exhibited the lowest R-Squared at 74%, as shown in Table VII. The Random Forest Regressor model within the SMRM hybrid system demonstrated the ability to adapt to the data, meaning that the difference between the predicted and observed values is important (RMSE and MAE), as shown in Tables VIII and IX.

TABLE V. MAPE BY MULTIPLE REGRESSION MODELS

Algorithm	MAPE
KNN	-0,023
Random Forest	-0,021
SVM	-0,031
XGBoost	-0,022
Gradient Boosting	-0,022
Neural Network	-0,035
Decision Tree	-0,03
Ridge	-0,035
Lasso	-0,028
Elastic Net	-0,032
SMRM	-0,023

TABLE VI. MAPE BY INDIVIDUAL REGRESSION MODELS

Algorithm	MAPE
KNN	-0,023
Random Forest	-0,021
SVM	-0,031
XGBoost	-0,022
Gradient Boosting	-0,022
Neural Network	-0,035
Decision Tree	-0,03
Ridge	-0,035
Lasso	-0,028
Elastic Net	-0,032

TABLE VII. R2 BY MULTIPLE REGRESSION MODELS

Algorithm	R2
KNN	99,30%
Random Forest	99,40%
SVM	98,50%
XGBoost	99,40%
Gradient Boosting	99,40%
Neural Network	74,10%
Decision Tree	99%
Ridge	74%
Lasso	94,80%
Elastic Net	86,50%
SMRM	98,90%

TABLE VIII. MAPE BY MULTIPLE REGRESSION MODELS

Algorithm	RMSE
KNN	-2,043
Random Forest	-1,997
SVM	-3,162
XGBoost	-2,028
Gradient Boosting	-1,996
Neural Network	-7,721
Decision Tree	-2,538
Ridge	-7,725
Lasso	-4,351
Elastic Net	-6,071

TABLE IX. MAPE BY INDIVIDUAL REGRESSION MODELS

Algorithm	MAE
KNN	-1,303
Random Forest	-1,25
SVM	-1,77
XGBoost	-1,258
Gradient Boosting	-1,249
Neural Network	-1,428
Decision Tree	-1,68
Ridge	-1,42
Lasso	-1,39
Elastic Net	-1,43

## VI. CONCLUSION AND FUTURE WORK

Predicting crude oil prices is a difficult task that requires a nuanced understanding of a wide range of economic and political factors. The SMRM hybrid system proposed in this

paper is an innovative approach that leverages machine learning to better capture the complex relationships between these factors and crude oil prices. By combining multiple models, SMRM can generate more accurate and reliable predictions, which can help investors and traders make more informed decisions. The experiment results illustrate that SMRM surpasses existing prediction models in terms of both accuracy and stability, making it a valuable tool for anyone interested in predicting crude oil prices. While there is still much work to be done to fully understand and predict stock market trends, SMRM represents a major step forward in this field, and holds the potential to revolutionize the approach to crude oil price prediction in the coming years. Predicting crude oil prices is a complex and challenging task, requiring a nuanced understanding of economic and political factors. The proposed SMRM hybrid system represents a significant improvement over existing approaches, as it can leverage both quantitative and qualitative factors to generate more accurate predictions. By learning from past data, the system can continually improve its forecasts, making it a robust and flexible forecasting tool that can support decision-making for a range of stakeholders, including investors and policymakers. Experiments demonstrate that SMRM outperforms existing models in terms of accuracy and stability, highlighting its potential as a powerful tool for predicting crude oil prices in the years to come. However, the proposed SMRM hybrid system offers a robust tool for predicting crude oil prices with heightened accuracy and reliability. However, there are still challenges to address, such as quantifying the impact of irregular factors like political risks and extreme weather events on crude oil prices. To address these challenges, future research will aim to incorporate these factors into the SMRM hybrid system and quantify their impact, leading to even more accurate predictions. With continued research and development, SMRM has the potential to revolutionize crude oil price prediction and help stakeholders make more informed decisions in the dynamic and complex world of stock market trading. In conclusion, the proposed SMRM hybrid system offers a promising solution for predicting crude oil prices, leveraging the power of machine learning, and combining multiple models to better capture the complex relationships between different factors. The experiments reveal that SMRM excels over existing models in both accuracy and stability, making it a valuable tool for investors, traders, and other stakeholders in the energy sector. The system can also be continually refined and improved by incorporating irregular factors like political risks and extreme weather events, which can help to better predict changes in crude oil prices. With further development, this approach could have important implications for supporting decision-making and risk management in the energy sector, enabling stakeholders to make more informed and effective decisions in the dynamic and complex world of stock market trading. By providing more accurate and reliable predictions of crude oil prices, the SMRM hybrid system has the potential to revolutionize how approach to predicting crude oil prices, providing valuable insights that can help to optimize decision-making and drive greater value in the energy sector.

REFERENCES

- [1] Khashei, M., & Mahdavi Sharif, B. (2021). A Kalman filter-based hybridization model of statistical and intelligent approaches for exchange rate forecasting. *Journal of Modelling in Management*, 16(2), 579-601.
- [2] C. Hamzacebi, "Improving artificial neural networks' performance in seasonal time series forecasting", *Information Sciences* 178 (2008), pages: 4550-4559.
- [3] G.P. Zhang, "A neural network ensemble method with jittered training data for time series forecasting", *Information Sciences* 177 (2007), pages: 5329-5346.
- [4] G.P. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model", *Neurocomputing* 50 (2003), pages: 159-175.
- [5] H. Park, "Forecasting Three-Month Treasury Bills Using ARIMA and GARCH Models", *Econ* 930, Department of Economics, Kansas State University, 1999.
- [6] L.J. Cao and Francis E.H. Tay "Support Vector Machine with Adaptive Parameters in Financial Time Series Forecasting", *IEEE Transaction on Neural Networks*, Vol. 14, No. 6, November 2003, pages: 1506-1518.
- [7] R. Lombardo, J. Flaherty, "Modelling Private New Housing Starts In Australia", *Pacific-Rim Real Estate Society Conference*, University of Technology Sydney (UTS), January 24-27, 2000.
- [8] Ahmad, A., Javaid, N., Guizani, M., Alrajeh, N., & Khan, Z. A. (2016). An accurate and fast converging short-term load forecasting model for industrial applications in a smart grid. *IEEE Transactions on Industrial Informatics*, 13(5), 2587-2596.
- [9] Yu, L., Wang, S., & Lai, K. K. (2008). Forecasting crude oil price with an EMD-based neural network ensemble learning paradigm. *Energy Economics*, 30(5), 2623-2635.
- [10] Xiong, T., Bao, Y., & Hu, Z. (2013). Beyond one-step-ahead forecasting: evaluation of alternative multi-step-ahead forecasting models for crude oil prices. *Energy Economics*, 40, 405-415.
- [11] Kumar, M. S. (1992). The forecasting accuracy of crude oil futures prices. *Staff Papers*, 39(2), 432-461.
- [12] Liu, Jinlan, Yin Bai, and Bin Li. "A new approach to forecast crude oil price based on fuzzy neural network." *Fuzzy Systems and Knowledge Discovery*, 2007. FSKD 2007. Fourth International Conference on. Vol. 3. IEEE, 2007.
- [13] Alizadeh, A., and Kh Mafinezhad. "Monthly Brent oil price forecasting using artificial neural networks and a crisis index." *Electronics and Information Engineering (ICEIE)*, 2010 International Conference On. Vol. 2. IEEE, 2010.
- [14] Safari, A., & Davallou, M. (2018). Oil price forecasting using a hybrid model. *Energy*, 148, 49-58.
- [15] Yi, Yao, and Ni Qin. "Oil price forecasting based on selforganizing data mining." *Grey Systems and Intelligent Services*, 2009. GSIS 2009. IEEE International Conference on. IEEE, 2009.
- [16] Nwulu, N. I. (2017, September). A decision trees approach to oil price prediction. In *2017 International Artificial Intelligence and Data Processing Symposium (IDAP)* (pp. 1-5). IEEE.
- [17] Khaidem, L., Saha, S., & Dey, S. R. (2016). Predicting the direction of stock market prices using random forest. *arXiv preprint arXiv:1605.00003*.
- [18] Tang, L., Pan, H., & Yao, Y. (2018, March). K-Nearest Neighbor Regression with Principal Component Analysis for Financial Time Series Prediction. In *Proceedings of the 2018 International Conference on Computing and Artificial Intelligence* (pp. 127-131).
- [19] Tang, L., Pan, H., & Yao, Y. (2018). PANK-A financial time series prediction model integrating principal component analysis, affinity propagation clustering and nested k-nearest neighbor regression. *Journal of Interdisciplinary Mathematics*, 21(3), 717-728
- [20] Zhang, Y., He, J., & Yin, T. F. (2012). Research on petroleum price prediction based on SVM. *Computer Simulation*, 29(3), 375.
- [21] Yu, L., Zhang, X., & Wang, S. (2017). Assessing potentiality of support vector machine method in crude oil price forecasting. *Eurasia Journal of Mathematics, Science and Technology Education*, 13(12), 7893-7904.
- [22] Kumar, Y. J. N., Preetham, P., Varma, P. K., Rohith, P., & Kumar, P. D. (2020, July). Crude Oil Price Prediction Using Deep Learning. In *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)* (pp. 118-123). IEEE.
- [23] Wang, S., Yu, L., & Lai, K. K. (2004, July). A novel hybrid AI system framework for crude oil price forecasting. In *Chinese Academy of Sciences Symposium on Data Mining and Knowledge Management* (pp. 233-242). Springer, Berlin, Heidelberg.
- [24] Hafezi, R., Shahrabi, J., & Hadavandi, E. (2015). A bat-neural network multi-agent system (BNNMAS) for stock price prediction: Case study of DAX stock price. *Applied Soft Computing*, 29, 196-210.
- [25] Nguyen, H. V., Naeem, M. A., Wichitakorn, N., & Pears, R. (2019). A smart system for short-term price prediction using time series models. *Computers & Electrical Engineering*, 76, 339-352.
- [26] Cheng, Y., Yi, J., Yang, X., Lai, K. K., & Seco, L. (2022). A CEEMD-ARIMA-SVM model with structural breaks to forecast the crude oil prices linked with extreme events. *Soft Computing*, 26(17), 8537-8551.
- [27] Kaymak, Ö. Ö., & Kaymak, Y. (2022). Prediction of crude oil prices in COVID-19 outbreak using real data. *Chaos, Solitons & Fractals*, 158, 111990.
- [28] Wu, B., Wang, L., Wang, S., & Zeng, Y. R. (2021). Forecasting the US oil markets based on social media information during the COVID-19 pandemic. *Energy*, 226, 120403.
- [29] Shen, Z. (2022, July). Optimal Oil-based Exotic Options Strategies Under the Background of War: An Empirical Study in the Context of the Russia-Ukraine Conflict. In *2022 2nd International Conference on Enterprise Management and Economic Development (ICEMED 2022)* (pp. 954-961). Atlantis Press.
- [30] Sun, Y. (2022, July). The Impacts of Wars on Oil Prices. In *2022 3rd International Conference on Mental Health, Education and Human Development (MHEHD 2022)* (pp. 167-170). Atlantis Press.
- [31] Ha, L. T. (2022). Dynamic interlinkages between the crude oil and gold and stock during Russia-Ukraine War: evidence from an extended TVP-VAR analysis. *Environmental Science and Pollution Research*, 1-14.
- [32] Yuan, X., & Li, X. (2021). Mapping the technology diffusion of battery electric vehicle based on patent analysis: A perspective of global innovation systems. *Energy*, 222, 119897.
- [33] Wang, J., Zhou, H., Hong, T., Li, X., & Wang, S. (2020). A multi-granularity heterogeneous combination approach to crude oil price forecasting. *Energy Economics*, 91, 104790.
- [34] Vijh, M., Chandola, D., Tikkiwal, V. A., & Kumar, A. (2020). Stock closing price prediction using machine learning techniques. *Procedia computer science*, 167, 599-606.
- [35] Zhang, J., Li, D., & Wang, Y. (2020). Predicting uniaxial compressive strength of oil palm shell concrete using a hybrid artificial intelligence model. *Journal of Building Engineering*, 30, 101282.
- [36] Abdollahi, H. (2020). A novel hybrid model for forecasting crude oil price based on time series decomposition. *Applied energy*, 267, 115035.
- [37] Bristone, M., Prasad, R., & Abubakar, A. A. (2020). CPPCNDL: Crude oil price prediction using complex network and deep learning algorithms. *Petroleum*, 6(4), 353-361.
- [38] Abdollahi, H., & Ebrahimi, S. B. (2020). A new hybrid model for forecasting Brent crude oil price. *Energy*, 200, 117520.
- [39] Wang, J., Lei, C., & Guo, M. (2020). Daily natural gas price forecasting by a weighted hybrid data-driven model. *Journal of Petroleum Science and Engineering*, 192, 107240.
- [40] Yahoo Finance, <https://finance.yahoo.com/quote/CL%3DF/history?p=CL%3DF>

# Compression Analysis of Hybrid Model Based on Scalable WDR Method and CNN for ROI-based Medical Image Transmission

Dr. Bindulal T.S.

Assistant Professor, Dept. of Computer Science, Govt. College, Nedumangad, University of Kerala, India

**Abstract**—The image compression techniques are the fast-growing methods and have developed on large scale. Among them, wavelet-based compression methods are most promising and efficient techniques widely used in the field of medical image processing and transmission. The compression techniques are treated as lossy or lossless models and these can be applied on the medical images considering different situations. The medical image parts are separated into two regions. The central part of the image is treated as core region called region of interest (ROI) and others are treated as non-ROI. ROI based coding techniques are considered as most important in the medical field for efficient transmission of clinical data. The proposed method focuses on these concepts. The ROI parts considered are either smooth or textured regions. These are extracted using a segmentation method called singular value decomposition (SVD) method. An efficient run length coding method called wavelet difference reduction method (WDR) with region growing approach is used to code the extracted ROI part after applying 5/3 based integer wavelet transform. The remaining parts called non-ROI part or background artifacts are coded using Convolution Neural network (CNN) method. The proposed method is also restructured as layered structure to achieve adaptive scalable property and named as scalable WDR-CNN (SWDR-CNN) method. The proposed SWDR-CNN method has been evaluated using rate distortion metrics such as Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index (SSIM). The coding gains in terms of PSNR values of SWDR-CNN method has been analysed and compared to popular scalable algorithm like S-SPIHT. The SWDR-CNN method has achieved better coding gain from 0.2 dB to 6 dB in terms of PSNR values. Hence, it is proved that proposed model can be used to code the ROI of images and has applications in the field of medical image data coding and transmission.

**Keywords**—*Medical image segmentation; compression; region of interest; wavelet difference reduction; convolutional neural network; singular value decomposition*

## I. INTRODUCTION

Design and development of efficient image processing methods is a subject of intense research for the last several decades. The demand of properly designed algorithm is much more in image coding [1] and transmission scenario. Advanced development in the networking technologies [2] with high processing power and transmission capacity, make it possible to implement modern signal processing [3] techniques in data computing scenarios. Hence, new technologies are developed, combining more than one method called hybrid model [4] to address specific user requirements. Moreover, tremendous

growth of the Internet and data sharing facilities, the developed hybrid model addresses the situations to achieve better computing capabilities in different network access bandwidths.

The image coding algorithms are developed and widely accepted with multiresolution features which are based on the JPEG2000 image compression standard. For Human Visual Systems, multi-resolution based coding algorithms are developed by using [5] wavelet transform. Thus, wavelet transform is used as supportive framework for decomposing images with different time-scale resolution form. The data analysis in wavelet domain has significant impact on the medical image computing scenario like in disease diagnosis and visualization. Due to the powerful multiscale/multiresolution representation of data in wavelet domain, many wavelet-based coding algorithm are developed for magnetic resonance Imaging (MRI) technologies.

Considering the inter scale/intra scale properties of wavelet transform, different approaches were developed in the late 1990s and 2000s. Among them, Shapiro's [6] Embedded Zero-tree Wavelet (EZW) coding scheme, Set Partitioning in Hierarchical Trees (SPIHT) [7] introduced by Said and Pearlman are most important. Due to the spatial orientation tree (SOT) based algorithm, the computational complexity of these algorithms remained very high. Hence, alternate methods that avoid the heavy use of SOTs, without sacrificing the desired properties of embedded coding, progressive transmission and scalability were also required to be developed.

Considering the time complexity, a new method called Wavelet Difference Reduction (WDR) method [9, 10] was developed by Tian and Well. The spatial orientation tree-based data structure was precluded in this method, but preserves the embedded principles, lossless bit plane coding and set partitioning concepts. The wavelet coefficients are linearly arranged using a fixed scan path by mapping the 2-D transform coefficients to 1-D index array which are present in the multiresolution pyramidal structure of wavelet transform. The WDR performs the run between two neighbouring significant coefficients and takes the difference between their indices and then these indices are coded efficiently. The WDR also maintains the simplicity while keeping the coding performance advantages of spatial orientation tree-based methods like EZW and SPIHT. Further, WDR methods were improved by incorporating many other desired features [24, 25] like coding regions of interest, scalability [26, 27, 28], object-based shape coding etc. Among them, Adaptive Scanned WDR (ASWDR)

of Walker and Ngan [11] considers one level of parent-child relationship, embedded progressive coding method developed by Wilhelm Berghorn [22, 23] with context conditioning and Context Modelling with ASWDR method (CMWDR) developed by Yuan and Mandal [12] considered much better method with coding gain than standard SPIHT algorithm even if the absence of entropy coding.

Significant development of coding methods is progressed to incorporate scalability concepts to decode the data by different resolution capacity devices on particular bit rate. David Taubman proposed a method called EBCOT, originally based on JPEG2000 standard [31] which supports SNR scalability. These types of algorithms are suitable for applications like remote browsing of large compressed images. The scalable image transmission is possible by using the layered approach and Hwang and Chine [43] proposed a medical image compression and transmission called Layered SPIHT. In LSPHIT, the bit streams are generated from different layers of subbands. According to time scale property of wavelet sub-bands, subbands are arranged on the basis of the priority and the bit streams are produced progressively with scalable property. Astri Handayani [44] proposed a medical image transfer method considering the ROI over heterogenous network, where good image quality and data compactness are both of crucial concern. The method follows wavelet-based coding techniques with layered approach as used in JPEG2000.

In ROI based image compression scenario, the new technique called object coding was developed by using shape adaptive wavelet transform, proposed [18] by Shipeng Li in 2000. In 2005, Ping Xu [19] and Shan'an Zhu proposed a method for arbitrary shaped ROI coding based on integer wavelet transform. Using this arbitrary shaped adaptive wavelet transform, Danyali [8] proposed the flexible scalable object coding using SPIHT method. Mehrotra, Srikanth and Ramakrishnan [30] proposed their coding method for MR images using shape adaptive integer wavelet transforms.

Generally, the medical images are coded using lossless image compression techniques and that can be done using integer wavelet transform. Many of the methods with progressive transmission facility have poor rate distortion performance in low bit rates. In medical images, the central portion of the image is considered as most important part. These central parts are either smooth or textured regions and can be extracted using some segmentation techniques. These segments are coded with maximum priority in lossless manner. The background can be repeatedly coded with lossy/lossless algorithm. In this paper, the objects are identified as ROI using singular value decomposition method and are coded using scalable WDR method [13, 14, 15, 16, 17]. The scalable WDR method is a layered architecture of wavelet different reduction method with embedded region growing method (WDR-SRG). The lossless coding is performed by using arbitrary shaped integer wavelet transform and SWDR. The remaining parts are coded using lossy coding techniques. Here, we use convolution neural network (CNN) model to code non-ROI area of the image. In 2020, Chung K. J. et al. proposed [34] a cross-domain cascade of U-nets that was the W-net compression technique and operated over the discrete cosine transform (DCT). In 2020, Guo P. et al. [41] had introduced the CNN-

based compression technique. The model focuses on retina optical coherence tomography (OCT) images. The model was trained and tested on OCT images with pathological details.

Here, the paper presents a hybrid model which includes SVD based image segmentation and DWT based object coding. The layered region-growing approach [15] based WDR method is used in DWT Coding algorithm. Also, CNN-based medical image compression [35] technique with minimum information loss is used for non-ROI. The coding performance of proposed model is analysed in terms of PSNR values and SSIM [41] metric. The analysis shows that the coding gain is much better in terms of PSNR values from 0.2 dB to 6 dB in various bit rates than the traditional coding schemes like SPIHT and its scalable version.

## II. SEGMENTATION USING SVD METHOD

Segmentation is a process to identify important portion of an image as per the user requirements. Most probably, the central portion of the medical images is considered as ROI which consists of important information. Hence, ROI part should be extracted for lossless data coding and can be done very efficiently by using the SVD method. SVD method [15] is based on eigen value and eigen vector analysis and it is observed that a small Eigen values in the non-ROI part will be generated. Therefore, applying some optimum threshold value on Eigen value, we can extract the ROI part of the image. The eigen value analysis applied on image to get textured region is explained below.

Consider the image,  $I(x, y)$ . Before starting the SVD analysis, windowing method is applied on the image  $I(x, y)$  and collected fixed sized windows which are used for labelling content feature. The textured regions are identified in terms of corners and edges. These can be easily separated by doing the gradient calculation. Let  $L_i$  is the linearly arranged signal values in window 'w'.  $\nabla L_i$  is the gradient in w.

$$\nabla L(i) = \left( L_x(i), L_y(i) \right)^T \quad (1)$$

where  $L_x(i) = \partial L / \partial x$  and  $L_y(i) = \partial L / \partial y$

The autocorrelation matrix is calculated as follows:

$$C = \begin{bmatrix} \sum L_x^2(i) & \sum L_x(i)L_y(i) \\ \sum L_x(i)L_y(i) & \sum L_y^2(i) \end{bmatrix} \quad (2)$$

The generated 2x2 matrix C is as symmetric autocorrelation matrix and eigen value analysis is applied on this matrix. After analysis, we can write symmetric matrix C as,

$$C = UDU^T \quad (3)$$

where, U is the ortho-normal column vector and D is the diagonal matrix. The diagonal matrix consists of two eigen values. The matrix values satisfy the following condition such that  $diag(e_1, e_2)$ ,  $e_1 \geq e_2$ , where  $e_i$  are the eigen values of the autocorrelation matrix C. Considering the values of  $e$ 's, the segmentation can be done according to the following conditions.

1) Let the window w contains a smooth region, then corresponding eigen values  $e_1$  and  $e_2$  are small.

- 2) Let the window  $w$  contains edges or corners, then first eigen value  $e_1$  is of principal component and the second eigen value  $e_2$  is of extremely small magnitude.
- 3) Let the content in window  $w$  consists of textures and patterns, then eigen value  $e_2$  is significant one.

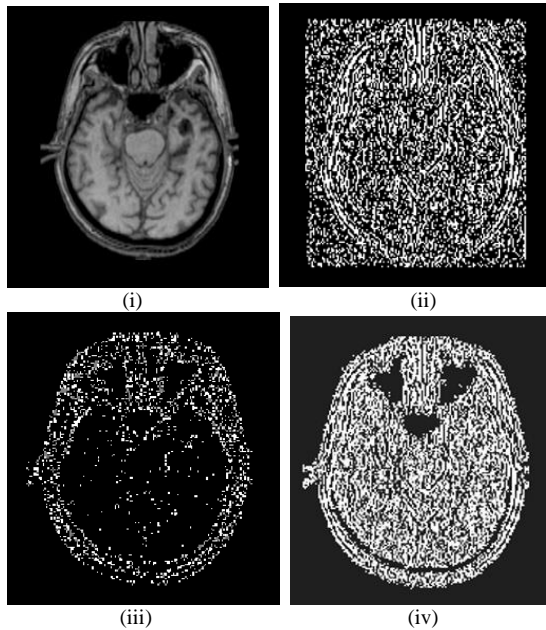


Fig. 1. Segmentation using SVD (i) Original MRI image (512x512) (ii) binary representation (iii) identified ROI (T = 1000 for eigen value) (iv) identified ROI (T = 10 for eigen value).

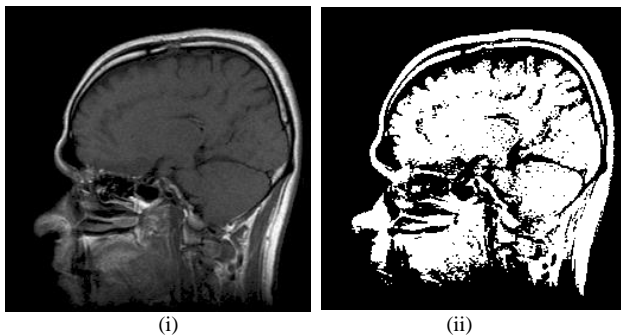


Fig. 2. Segmentation using SVD (i) Original MRI image (256x256) (ii) identified ROI.

The texture extraction is done by applying threshold value on eigen values ( $e_1, e_2$ ) satisfying the above three conditions. Segmentation using SVD analysis of *MRI* medical image is shown in Fig. 1 and Fig. 2. The eigen values are calculated on each block with window size  $2 \times 2$ . The obtained eigen values are from 0 to a value of power of 10. A threshold value  $T$  can be applied to remove negligible correlation of values and it may be in noise part or smooth region of the image. The obtained eigen values from 100 window blocks are shown in the Fig. 3.

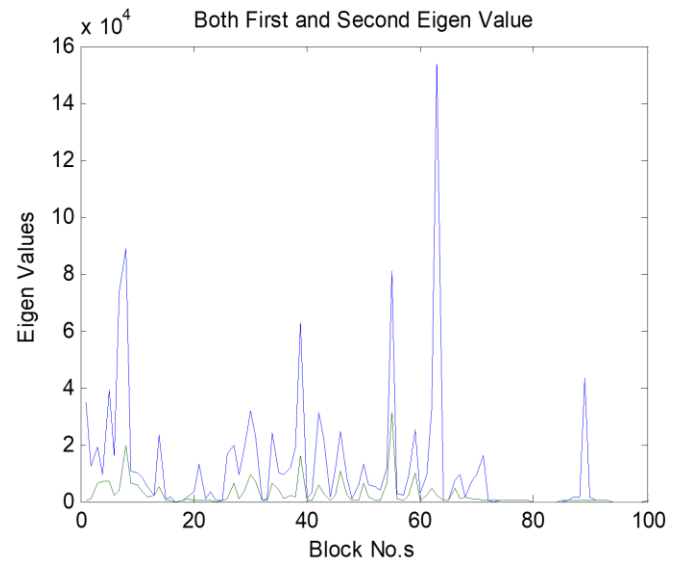


Fig. 3. Eigen values of first 100 blocks (window -  $2 \times 2$ ).

### III. LAYERED SCALABLE WDR METHOD WITH SELECTED REGION GROWING APPROACH (SWDR-SRG)

The ROI based coding is important in many applications in telemedicine so that important medical data needs to be coded without loss. The balance between coding of ROI and BG of the medical images is maintained by using object coding. Moreover, Integer Wavelet Transform is used to improve the quality of the reconstructed ROI. The Shape Adaptive DWT is used in object coding and also maintains spatial correlation between the actual data and its transformed data. Moreover, it keeps count of the coefficients obtained from SA-DWT [18, 19] is same to the count of pixels in the region.

The shape adaptive DWT [20, 29, 30] is done based on the procedure as follows. The process begins by identifying arbitrary shaped objects using a segmentation method. The first segment of the object is applied by length adaptive 1-D DWT with proper subsampling. The calculated wavelet coefficients are categorized into low pass and high pass bands. After the completion of row wise operation, each column of the low pass and high pass objects are applied by the same operations. The same operation is repeatedly applied on object in low pass band to get desired level of wavelet decomposition. Thus, 2D SA-DWT provides multi-resolution pyramid of arbitrary shaped objects. The subband structure of arbitrary shaped object is shown in the Fig. 4.

The segmentation process is carried out as a preprocessing in image compression model. The main aim is not only object identification, but classifying spatially connected homogeneous pixels present in a small region with similar gray levels. In this situation, region growing approach is better for ROI based image coding. The Selected Region Growing starts with a seed pixel and connected with four neighbouring pixels with similar features. The process progresses until last seed pixel reaches and terminated when condition fails. Here, the SRG process is done on the transform values which are present in the wavelet domain. The significant coefficients are identified in the extracted region. WDR method is used here to perform SRG operation. Index coding with differential coding method is used

in WDR method where it codes the index positions of significant transform values very efficiently. The 2D wavelet transform coefficients are linearly arranged in increasing order of the index positions using a predefined scan path [21]. The general structure of algorithm WDR-Selected Region Growing (WDR-SRG) [15] is depicted below:

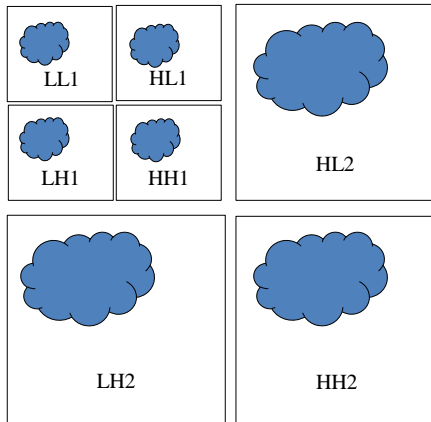


Fig. 4. Subband structure of objects using shape adaptive DWT.

#### A. Wavelet Difference Reduction with Selected Region Growing (WDR-SRG)

Algorithm starts with two-dimensional shape adaptive DWT which is applied on the arbitrary shaped objects. The decomposition of the medical ROI is done using a 5/3-tap wavelet filter with symmetric extension. Wavelet coefficient at pixel location  $(i, j)$  is represented as  $w_{ij}$ . The iterative operation is progressed within the limit of a threshold value and initialised as,  $t_{n-1} = 2^{n+1}$  and  $t_n = t_{n-1}/2$ , where  $n = \left\lfloor \log_2 \left( \max_{(i,j) \in I} |w_{ij}| \right) \right\rfloor$ . The algorithm is progressed through sorting pass and refinement pass. These are narrated as below.

1) *Sorting pass*: The sorting pass uses different data structures or lists to store the positions of coefficients which are collected during the process. The collection of coefficients during region growing process is stored in the list R, significant coefficients are stored in the list C and its neighbouring pixels are stored in the list N. The parent child connection pixels are stored in the list P. The list T is used to store temporary set of significant coefficients. Also, the list I is used to store set of remaining insignificant coefficients. The significant coefficients are identified by using  $\sigma(w, t_n)$  such as,

$$\sigma(w, t_n) = \begin{cases} 1 & : |w| \geq t_n \\ 0 & : |w| < t_n \end{cases} \quad (4)$$

The sign of coefficient 'w' is identified by using function  $Sign(w)$  such as,

$$Sign(w) = \begin{cases} + & : w \geq 0 \\ - & : w < 0 \end{cases} \quad (5)$$

The  $cluster(w_{ij}, t_n)$  function is to collect the neighbouring coefficients of the significant coefficients. The sorting pass procedure is depicted below:

```

If I ≠ ∅ {
    If σ(wij, tn-1) = 0 {
        If σ(wij, tn) = 1 {
            Coding (wij);
            R = cluster(wij, tn);
        }
        Do {
            If R ≠ ∅ {
                If σ(R(wij, tn)) = 1 {
                    Coding ( wij )
                    R = cluster(wij, tn)}
            } While (End (R) ≠ True);
        }
    }
}
    
```

The encoding process is done by differential coding using a function Coding ( $w_{ij}$ ). The function procedure is depicted below:

#### Function Coding ( $w_{ij}$ )

```

{
    Binary representation of value obtained as distance between
    two significant coefficients avoiding MSB '1'. The sign
    information generated using the function Sign(wij). Append
    wij into temporary list T.
}
    
```

The next step of sorting pass is progressed only after the updation of list and that is depicted in the Index updating pass as shown below:

```

If T ≠ ∅ {
    N = cluster(wij, tn) ; ∀(i, j) ∈ T
    P = child(wij, tn) ; ∀(i, j) ∈ T
}
    
```

List of insignificant coefficients, I is reset after removing these coefficients

$$I = R + N + P + I.$$

2) *Refinement pass*: After the sorting pass, the collected significant coefficients are coded as bit plane coding manner using the refinement pass as shown below:

```

If C ≠ ∅ {
    If C(σ(wij, tn-1) = 1) {
        Add nth MSB of C(wij).
    }
    C = C + T.
    T = ∅
}
    
```

Next iteration starts after the threshold update process. i.e.,  $t_{n-1} = t_n$ ,  $t_n = t_n/2$ . The encoding procedure generates four symbols like +, -, 1 and 0. The entropy coding like arithmetic coding is avoided by replacing the symbols by using two bits like 11 for +, 10 for -, 01 for 1 and 00 for 0.

### B. Scalable Wavelet Difference Reduction Method

The developed coding scheme WDR-SRG is quite useful in the field of communication over heterogeneous networks if it supports the concepts of adaptive scalability. For that, wavelet difference reduction method with selected region growing (WDR-SRG) is applied on ROI after the SA-DWT. Thus, the base algorithm is restructured in layered manner so that the algorithm can generate different scalable images. Hence, we renamed the method as scalable WDR (SWDR) [13, 14, 16] method. Thus, the coding procedure is following the adaptive scalable concepts. The coding procedure is depicted in Fig. 5. The scalable image generation from the layered bit stream is depicted in Fig. 6.

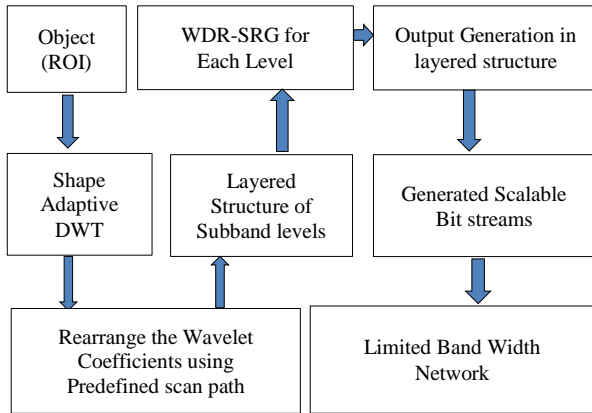


Fig. 5. Scalable WDR model (SWDR).

### C. Algorithm: ROI coding using Scalable Wavelet Difference Reduction Method

Step 1: The MRI Medical Image is inputted in the procedure,  $I(x, y)$

Step 2: The SVD is applied on MRI medical image to identify the ROI

- $I(x, y) \rightarrow O(x, y) + N(x, y)$
- Input Image = Object Image + Noise
- Input Image = ROI + nonROI

Step 3: The ROI is treated as object and applied SA-DWT

- $O(x, y) \rightarrow SA(x, y)$

Step 4: The Layered Structure of Scalable WDR is applied.

- $SA(x, y) \rightarrow SWDR(x, y)$

Step 5: The bit stream is generated for transmission

- $SWDR(x, y) \rightarrow E(Binary)$

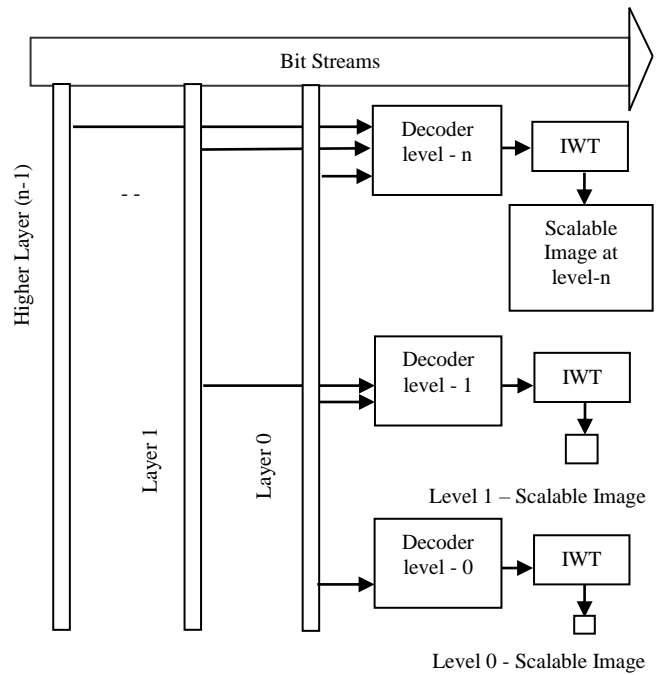


Fig. 6. Scalable image generation (layered architecture).

### IV. NON-ROI CODING USING CNN MODEL

During the image compression process, the artifacts are identified with several features. These artifacts may be suppressed efficiently by using CNN model. In this field, many of the scientists [32, 33, 34] have developed efficient CNN models which can be used in data coding situations. Fully convolutional model is also used to obtain a compact representation of an image. The general idea of CNN model is shown in Fig. 7. The image features are identified and represent as parameters to train the data how it can be represented in compact form. Here, a series of convolutional layers are fixed in such a way that features of [35, 36, 37] images are captured. Thus, structural configuration of an image is maintained as well. The model shown in Fig. 7 describes the CNN Forward as a well-designed network [38, 39, 40] which is used to represent the image in identified properties. Thus, CNN forward is used to compress in such a way that original data can be reproduced by reconstruction network. The model consists of three convolutional layers in which the second layer followed by batch normalization layer. After the operation of first convolutional layer, the image size is reduced by half. The data in the form of grid is used in Convolutional Neural Networks. In CNNs, each kernel produces in a new convoluted layer. These layers are also called activation maps. In CNN model, the convolution operation is used such that a kernel or a mask moves over an image and a convoluted representation of output is generated [40, 41]. The identified dependencies in an image are captured by CNN.

CNN Model [45] with different layers is used to perform convolution operation on matrix data. The image is inputted as matrix with two dimensional or three-dimensional form. The other matrix called filter matrix or kernel matrix is inputted with features like height, width and dimension. Let inputted image is  $I(x, y)$ , then filter matrix is  $F(h, w, d)$ , where  $h$  is



height,  $w$  is width and  $d$  is dimension. After convolution operation, generated output matrix is  $C(x-h, y-w)$  for gray scale image. The basic architecture of CNN is shown in Fig. 8.

A. Algorithm: Non-ROI Coding using CNN Model

- Step 1. Input nonROI image  $\leftarrow N(x, y)$
- Step 2. Read  $N(x, y)$  and transform to CNN forward
- Step 3. ReLU based Two-dimensional convolution with Layer 1 and Layer 2  
Optimum Value  $\leftarrow$  Padding and striding is used in Max pooling two-dimensional layer 1
- Step 4. ReLU based Two-dimensional convolution with Layer 3 and Layer 4  
Optimum Value  $\leftarrow$  Padding and striding is used in Max pooling two-dimensional layer 2
- Step 5. ReLU based Two-dimensional convolution with Layer 5
- Step 6. NE(Binary)  $\leftarrow$  Compressed nonROI MRI medical image

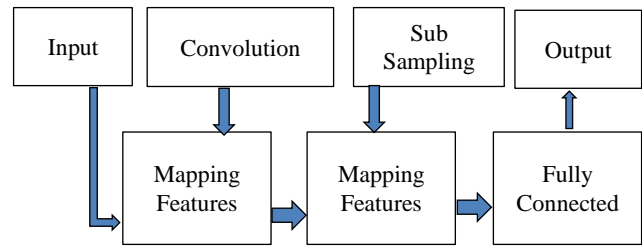


Fig. 8. Basic architecture of convolutional neural network.

The CNN model consists of three layers: convolution layers, pooling layers or subsampling, and fully connected layers. The model uses specific filters for image compression and decompression. The filter helps to represent the image in definite manner so that it removes the unnecessary features. The size of Kernel can be adjusted to increase the performance of data coding which is used in convolution operation at every point. The padding process is applied on all sides of image matrix to reduce the loss of data. The pooling layer is one layer used to reduce the feature map's dimension. But it is possible to recall necessary information. The optimum value is taken as maximum value which is done by Max pooling operation. Famous method called ReLU (Rectified Linear Unit) [32, 36, 37, 38] activation function is used in CNN for a non-linear operation. This method activates neurons in which the gradient provides all times the optimum value.

V. PROPOSED METHOD: HYBRID SCHEME WITH SCALABLE WDR AND CNN

The paper proposes a hybrid scheme with two coding techniques, one for the ROI and another for unimportant parts nonROI of the medical image. ROI parts are identified as object and coded using DWT based coding method (suitable for lossless image compression). The remaining nonROI parts called BG part is coded using CNN (useful for lossy image compression scheme). Medical image is inputted and SVD analysis is done for object identification. The shape adaptive DWT is applied on arbitrary shaped object and Scalable WDR is used to perform the lossless coding. The remaining BG part is coded using a lossy compression with CNN model. The outline of proposed method is shown in Fig. 9.

VI. EXPERIMENTAL RESULTS

This paper presents the compression analysis of a new fast hybrid method SWDR-CNN for medical image transmission. The MRI medical images are used for the simulation and analysis. The proposed scheme SWDR-CNN model is analysed in terms of the basic parameters like PSNR and SSIM. The hybrid model is useful for both lossy and lossless image coding. The quality of images at any bit rates is calculated as the peak signal to noise ratio (PSNR). It is defined as,

$$PSNR = 10 \log_{10} \left( \frac{max^2}{MSE} \right) \text{ dB} \quad (6)$$

where, MSE is mean squared error obtained by comparing the inputted image and the recreated image;  $max$  is the extreme value of a pixel inside the image.

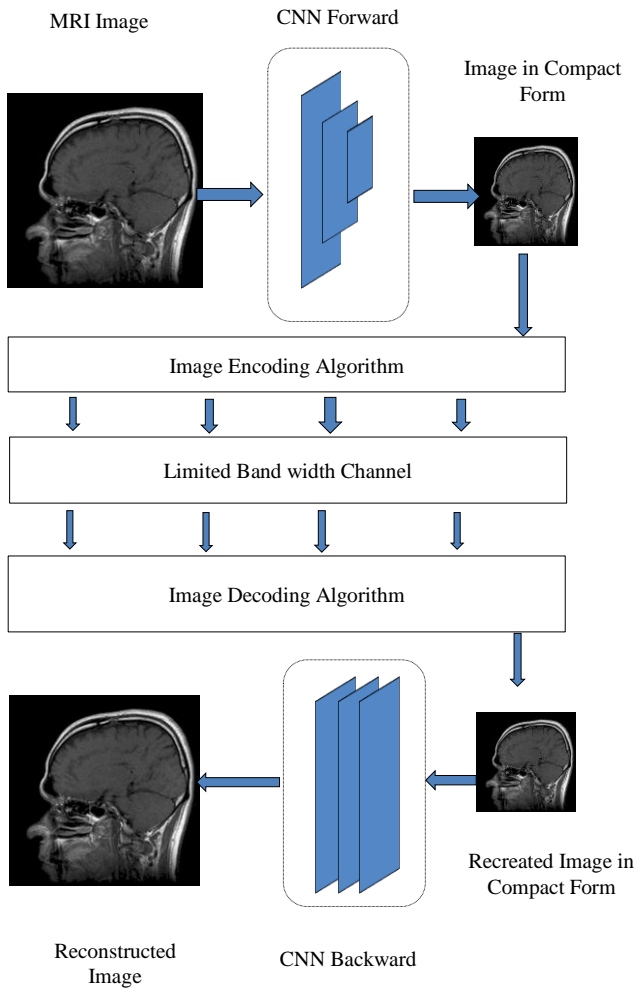


Fig. 7. CNN model for image compression.

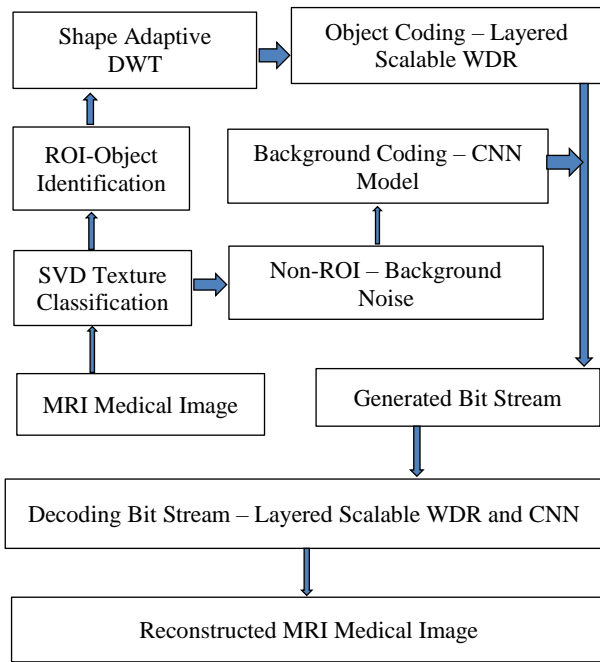


Fig. 9. Proposed compression method – SWDR-CNN.

Structural similarity index (SSIM) [42] is also used to analyse the performance used as rate distortion metric and can be defined in following equations. The SSIM between signals  $x$  and  $y$  is calculated as:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (7)$$

where  $\mu_x$  is mean of  $x$ ,  $\mu_y$  is mean of  $y$ ,  $\sigma_x^2$  is variance of  $x$ ,  $\sigma_y^2$  is variance of  $y$  and  $\sigma_{xy}$  is the covariance of  $x$  and  $y$ ,  $C_1$  and  $C_2$  are constants, i.e.,  $C_1 = (K_1L)^2$  and  $C_2 = (K_2L)^2$  where  $L$  is the dynamic range of the pixel values, and  $K_1$  and  $K_2$  are two constants. The overall quality value called the average of the quality map or the Mean SSIM (MSSIM) index is defined as,

$$MSSIM(X, Y) = \frac{1}{M} \sum_{j=1}^M SSIM(x_j, y_j) \quad (8)$$

where  $X$  is reference image and  $Y$  is the distorted image.  $x_j$  and  $y_j$  are the image contents at the  $j^{th}$  local window and  $M$  is the number of local windows of the image.

The rate distortion metric PSNR is calculated and coding gain of SWDR-CNN method is compared with the traditional methods like SPIHT and CMWDR without any entropy coding. Coding results for MRI test image with size 256x256 are shown in Table I and Fig. 10. Reconstructed images with size 256x256 are shown in Fig. 11. Coding results for MRI test image with size 512x512 are shown in Table II and Fig. 12. Reconstructed images with size 512x512 are shown in Fig. 13. The wavelet transform with 6-level is done for 512x512 images and 5-level is done for 256x256 images.

TABLE I. PSNR OF MRI IMAGE (256x256)

Bits per pixel	SPIHT	CMWDR	SWDR-CNN
0.125	26.9010	27.0121	27.1157

0.25	29.8901	30.0574	30.2246
0.5	33.1537	33.4325	33.6975
1.0	37.9891	38.1208	38.2972

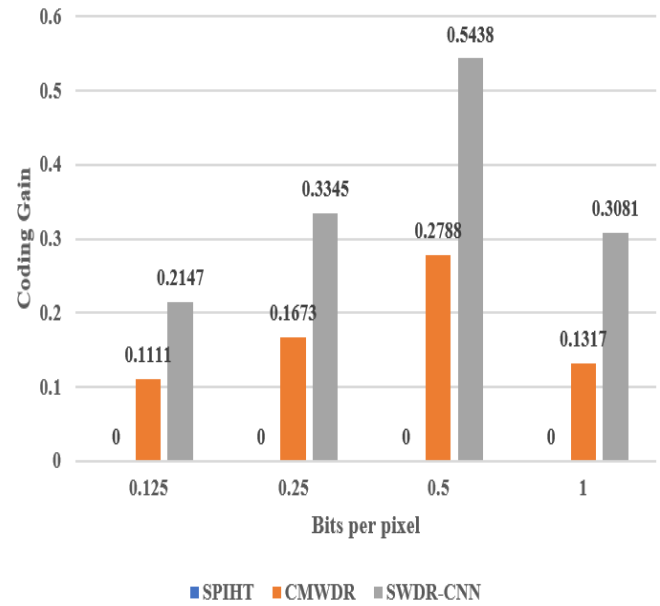
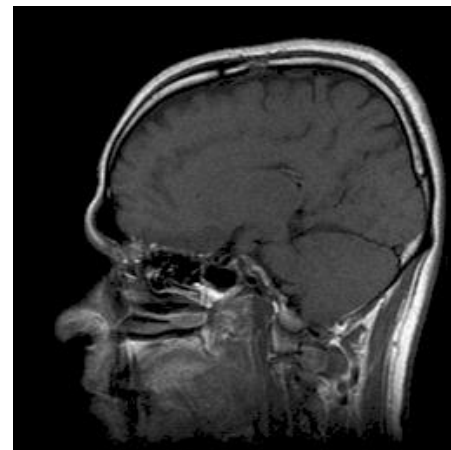


Fig. 10. Coding Gain (in dB) MRI Image (256x256).

TABLE II. PSNR OF MRI IMAGE (512x512)

Bit rate	SPIHT	CMWDR	SWDR-CNN
0.125	33.9175	34.1120	34.6784
0.25	38.1521	38.2831	38.9485
0.5	43.1421	43.5563	43.9025
1.0	49.1461	49.1145	49.5936



(i)

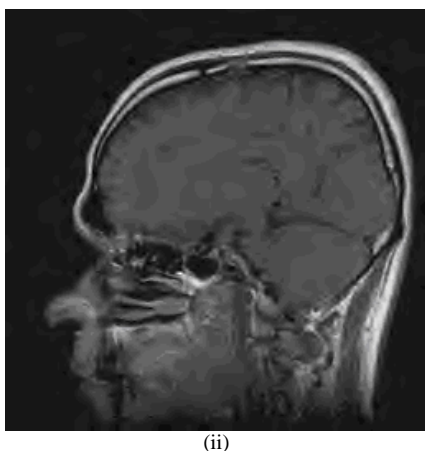


Fig. 11. Image reconstruction (bpp = 0.25). (i) original image (256x256) (ii)SWDR-CNN

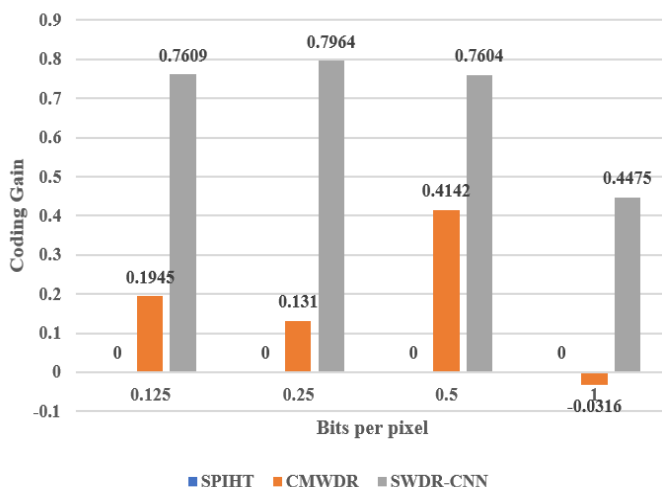


Fig. 12. Coding gain in dB of MRI image (512x512).

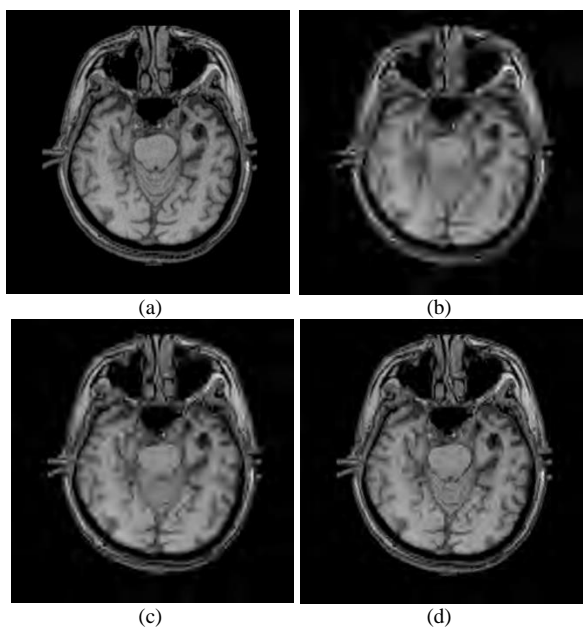


Fig. 13. Image reconstruction using SWDR-CNN Model at different bit rate (a) original image (512x512 - scale 40%) (b) 0.015625 (c) 0.03125 (d) 0.0625.

Structural Similarity Index (SSIM) of MRI image is shown in Table III. The proposed method is also simulated by incorporating with the scalability concept. Here, the comparison of the proposed model SWDR-CNN is done with layered form SPIHT method. Simulation results obtained are given in Table IV. For MRI with Full resolution, the coding is from 0.3 dB to 0.7 dB for different bpp. During the analysis, we identified one important point is that the coding gain in PSNR values (in dB) tremendously changes when the resolution scale decreases. At level 2 resolution 256 x 256, the coding gain is from 0.80 dB to 5.50 dB compared to SPIHT and from 0.7 dB to 1.5 dB compared to its scalable version at different bpp.

MATLAB tool in Windows OS is used to implement all the algorithm modules. The convolution-based DWT is developed using Bi-orthogonal 5/3 tap filter coefficients.

TABLE III. STRUCTURAL SIMILARITY INDEX VALUE OF MRI IMAGE (512x512)

Test Image	Bit rate (bpp)	Image Size (512x512)	
		SPIHT	SWDR-CNN
MRI	0.125	0.91672	0.91946
	0.25	0.95834	0.96079
	0.5	0.98360	0.98604
	0.75	0.99225	0.99297
	1	0.99605	0.99648

TABLE IV. PSNR VALUES IN DIFFERENT SCALABLE RESOLUTIONS

Level 1 - Full Resolution (512x512)				
Methods	Bits Per Pixel			
	0.125	0.25	0.5	1.0
SPIHT	-	-	-	-
Scalable SPIHT	31.11	35.79	40.31	47.05
SWDR-CNN	31.83	36.39	40.89	47.35
Level 2 - Half Resolution (256x256)				
Methods	Bits Per Pixel			
	0.0625	0.125	0.25	0.5
SPIHT	27.47	31.15	36.39	43.23
Scalable SPIHT	27.71	31.09	38.91	47.34
SWDR-CNN	28.49	32.05	39.52	48.71

## VII. CONCLUSION

The paper presents a hybrid scheme based on wavelet based scalable encoder with convolution neural network for image transmission. The scheme has high encoding performance with scalability property and can be used in medical image transmission field. The paper focused on efficient transmission of important parts present in medical images where quality is effectively preserved. It is required to review large images which are often required in the field of medical image computing. Moreover, data transfer is required over limited bandwidth channel in very low bit rate. Hence, object coding is essential for data transfer and here a modified

WDR called SWDR is used. These are implemented by using the arbitrary shaped DWT. Thus, new method has interesting perceptions for numerous medical data computing and transmitting applications. The medical data is classified into ROI and non-ROI with the help of SVD method. ROI parts are compressed using the SWDR and non-ROI parts are compressed using CNN model. The Shape Adaptive DWT in association with scalable WDR method is used as lossless compression technique and it is applied on an ROI portion of medical image data. Then, CNN model is applied on non-ROI as lossy coding scheme which is used to reduce the losses using weight and learning rate. After analysis, the projected method is good one with better coding speed with 12% gain than that of traditional methods. Moreover, better coding gain is obtained in terms of PSNR value from 0.2 dB to 6 dB in all situations. Thus, proposed coding scheme called SWDR-CNN model have better results than the existing coding models.

#### REFERENCES

- [1] Rafael C. Gonzalez, Richard E. Woods, "Digital Image Processing-Second Edition", Pearson Education, 2007.
- [2] Milan Sonka, Vaclav Itlavac, "Image Processing, Analysis and Machine Vision", PWS Publishing, 1998.
- [3] Al Bovik, "Hand book of Image and video processing", 2<sup>nd</sup> Edition, Academic Press, 2005.
- [4] David Salomon, "Data Compression, The Complete Reference", 3<sup>rd</sup> Edition, Springer, 2004.
- [5] M. Vetterli, J. Kovacevic, 'Wavelets and Subband Coding', Prentice Hall P T R, 1995.
- [6] J.M.Shapiro, "Embedded image coding using zero trees of wavelets coefficients", IEEE Trans. Signal Process. vol. 41, 1993, pp. 3445-3462.
- [7] A. Said, W. A. Pearlman, "A new fast and efficient image codec based on set partitioning in hierarchical trees", IEEE Trans. Circuits Syst. Video Technol. vol. 6, no.3, 1996, pp. 243-250.
- [8] H. Danyali, A. Mertins, "Flexible, highly scalable, object based wavelet image compression algorithm for network applications", IEE proceedings Vis. Image Signal Process., vol 151, No. 6, 2004, p.p.- 498-510.
- [9] J. Tian, Jr. R. O. Wells, "A lossy Image Codec Based on Index Coding", Proc. Data Compression Conference, IEEE Computer Society Press, 1996.
- [10] J. Tian, Jr. R. O. Wells, "Embedded image coding using wavelet difference reduction", Wavelet image and video compression, P. Topiwala, Ed. Norwell, MA: Kluwer, 1998, pp. 289-302.
- [11] James S. Walker, Truong Q. Nguyen, "Lossy image codec based on adaptively scanned wavelet difference reduction", Optical Engineering 39, 2000, pp. 1891-1897.
- [12] Yufei Yuan, Mrinal K. Mandal, "Novel embedded image coding algorithms based on wavelet difference reduction", In Proc. of IEE, Vol. 152, No. 1, Feb. 2005, pp. 9 – 19.
- [13] Bindulal T.S., M.R. Kaimal, "Adaptive Scalable Wavelet Difference Reduction Method for Efficient Medical Image Transmission", Proc. of IEEE, TENCON, Hong Kong, 2006.
- [14] Bindulal T.S, M.R. Kaimal, "Adaptive Scalable Wavelet Difference Reduction Method for Efficient Image Transmission", Lecture notes on Computer Science 4338, Springer, 2006, 708-717.
- [15] Bindulal T.S., M.R. Kaimal, "A Hybrid Scheme based on Wavelet Transform, SVD and WDR method for medical images", Proc. of IET International Conference on VIE 2006, India , p.p. 201-206.
- [16] Bindulal T.S., M.R. Kaimal, "Object coding using a shape adaptive wavelet transform with scalable WDR method", Proc. of IEEE International Conference on Image Processing (ICIP 2007), USA, Sept.-Oct. 2007, p.p. II-325-II-328.
- [17] Bindulal T. S, M. R. Kaimal, "Adaptive Coding Techniques for Efficient Image Processing", PhD Thesis, University of Kerala, December, 2009.
- [18] Shipeng Li, Weiping Li, "Shape adaptive Discrete Wavelet Transform for Arbitrarily Shaped Visual Object Coding", IEEE Trans. On Circuits Syst. Video Technol. Vol. 10, No. 5, 2006, pp. 725-43.
- [19] Ping Xu and Shannn Zhu, "A New Method for Arbitrarily Shape ROI Coding Based on ISA-DWT", Proc. Of IEEE Int. Conference on Control and Automation ICCA 2005, pp. 1018-21.
- [20] M. Caguazzo, G. Poggi, L. Verdoliva, "Costs and Advantages of Shape Adaptive Wavelet Transform for Region Based Image Coding", Proceedings of IEEE international Conference on Image Processing, Vol.3, 2005, pp. III-197-200
- [21] Zhexuan Song, Nick Roussopoulos, "Using Hilbert Curve in Image Storing and Retrieving", Proceeding of ACM Multimedia Wokshop, USA, pp.167-170, 2000.
- [22] Wilhelm Berghorn, "Fast Variable Run length coding for Embedded Progressive Wavelet Based image compression", IEEE Trans. On Image Processing, Vol. 10, No. 12, pp.1781-1790, 2001.
- [23] Wilhelm Berghorn, "Context Conditioning and Run length coding for Hybrid Embedded Progressive Image Coding", IEEE Trans. On Image Processing, Vol. 10, No. 12, pp.1791-1800, 2001.
- [24] Yee L. Law and Truong Q. Nguyen, "Motion Wavelet Difference Reduction (MWDR) Video Codec", Proc. Of IEEE Inter. Conference on Image Processing, pp.2303-2306, 2004.
- [25] Yee Louise Law, Frank Crosby, Quyen Huynh, Troung Nguyen, "Wavelet Difference Reduction with region of interest priority in multi-spectral video small target detection", Proc. Of International Conference on Image Processing, pp.1903-1906, 2004.
- [26] M. Marinov, D. Avresky, T. Nguyen, "Parallel and Reliable Execution of a WDR algorithm in high speed networks", Proc. Of IEEE International Conference and Workshops on the Engineering of Computer Based Systems, Computer Society, pp.27-32, 2005.
- [27] Yufei Yuan, Mrinal K. Mandal, "Embedded color Image Coding using Context modelled wavelet difference reduction", Proc. Of IEEE Inter. Conference on Acoustic Speech and Signal Processing, Vol. 3, pp. III-61-64, 2004.
- [28] Poonlap Lamsrichen, Teerapat Sanguankotchakorn, "Embedded Image Coding Using Context Based Adaptive Wavelet Difference Reduction", Proc. Of IEEE Inter. Conference on Image Processing, pp. 1137-1140, 2006.
- [29] Karl Martin, R. Lukac, K. N. Plataniotis, "SPIHT based Coding of the Shape and Texture of Arbitrarily Shaped Visual Objects", IEEE Trans. On Circuits and Systems in Video Technology, Vol.16, No.10, pp.1196-1208, 2006.
- [30] Mehrotra, R. Srikanth, A. G. Ramakrishanan, "A new Coding scheme for 2-D and 3-D MR images using shape adaptive integer wavelet transform", IEEE Conf. on Data Compression, pp.67-72, 2004.
- [31] Charilaos Christopoulos, Athenassios Skodras, "The JPEG2000 Still Image Coding System: An Overview", IEEE Trans. On Consumer Electronics, Vol. 40, No. 4, pp. 1103-1127, 2000.
- [32] Miao J., Sun K., Liao X., Leng L. & Chu, J. "Human Segmentation Based on Compressed Deep Convolutional Neural Network", IEEE Access, 2020, 8:167585-167595.
- [33] Tellez D., Litjens G., Laak J. & Ciompi F., "Neural image compression for gigapixel histopathology image analysis", IEEE trans. on pattern analysis & MI, 43(2), 2021.
- [34] Chung K. J., Souza R. & Frayne R. "Restoration of lossy JPEG-compressed brain MR images using cross-domain neural networks", IEEE Signal Proc. Letters, 2020, 27:141-145.
- [35] Sabbavarapu S. R., Gottapu S. R. & Bhima P. R., "A discrete wavelet transform and recurrent neural network based medical image compression for MRI and CT images", Journal of Ambient Intelligence and Humanized Computing, 2020, 1-13.
- [36] Rossinelli D., Fourestey G., Schmidt F., Busse B. & Kurtcuoglu V. "High-throughput lossy-to-lossless 3D image compression", IEEE Trans. on Med. Imaging, 2020, 40(2):607-620.

- [37] Zhou Y., Yen G. G. & Yi Z. "Evolutionary compression of deep neural networks for biomedical image segmentation", *IEEE trans. on N.N. and L. Sys.*, 2019, 31(8):2916-2929.
- [38] Nousias S., Arvanitis G., Lalos A. S., Pavlidis G., Koulamas C., Kalogeras A. and Moustakas K. "A saliency aware CNN-Based 3D Model simplification and compression framework for remote inspection of heritage sites", *IEEE Access*, 2020, 8:169982-170001.
- [39] Mardani M., Gong E., Cheng J. Y., Vasawala S. S., Zaharchuk G., Xing L. & Pauly J. M. "Deep generative adversarial neural networks for compressive sensing MRI", *IEEE transactions on medical imaging*, 2018, 38(1):167-179.
- [40] Devadoss C. P. & Sankaragomathi B. "Near lossless medical image compression using block BWT-MTF and hybrid fractal compression techniques", *Cluster Computing*, 2019, 22:12929-12937.
- [41] Guo P., Li D. & Li X. "Deep OCT image compression with convolutional neural networks", *Biomedical Optics Express*, 2020, 11(7):3543-3554.
- [42] Zhou, W., A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity", *IEEE Transactions on Image Processing*. 2004, Vol. 13(4), pp. 600-612.
- [43] Wen Jyi Hwang, Ching-Fung Chine, Kuo-Jung Li, "Scalable Medical Data Compression and Transmission Using Wavelet Transform for Telemedicine Applications", *IEEE Trans. On Information Technology in Biomedicine*, Vo. 7, No. 1, pp.54-63, 2003.
- [44] Astri Handayani, P. Rahmiati, A. B. Suksmono, T. L. R. Mengko, "Medical Image Transfer with Wavelet-based scalable Quality ROI Coding", *The Third APT Telemedicine Workshop*, 2005.
- [45] Raj Kumar Paul, Sravanan Chandran, "A health Care Image Compression Scheme using Discrete Wavelet Transform and Convolution Neural Network", *Journal of Engg. Research, ICMET Special Issue*, 2022.

# A Proposed Intelligent Model with Optimization Algorithm for Clustering Energy Consumption in Public Buildings

Ahmed Abdelaziz<sup>1</sup>, Vitor Santos<sup>2</sup>, Miguel Sales Dias<sup>3</sup>

Nova Information Management School, Universidade Nova de Lisboa, 1070-312 Lisbon, Portugal<sup>1,2</sup>  
Information System Department, Higher Technological Institute, HTI, Cairo 44629, Egypt<sup>1</sup>  
Instituto Universitário de Lisboa (ISCTE-IUL), ISTAR, 1649-026 Lisbon, Portugal<sup>3</sup>

**Abstract**—Recently, intelligent applications gained a significant role in the energy management of public buildings due to their ability to enhance energy consumption performance. Energy management of these buildings represents a big challenge due to their unexpected energy consumption characteristics and the deficiency of design guidelines for energy efficiency and sustainability solutions. Therefore, an analysis of energy consumption patterns in public buildings becomes necessary. This reveals the significance of understanding and classifying energy consumption patterns in these buildings. This study seeks to find the optimal intelligent technique for classifying energy consumption of public buildings into levels (e.g., low, medium, and high), find the critical factors that influence energy consumption, and finally, find the scientific rules (If-Then rules) to help decision-makers for determining the energy consumption level in each building. To achieve the objectives of this study, correlation coefficient analysis was used to determine critical factors that influence on energy consumption of public buildings; two intelligent models were used to determine the number of clusters of energy consumption patterns which are Self Organizing Map (SOM) and Batch-SOM based on Principal Component Analysis (PCA). SOM outperforms Batch-SOM in terms of quantization error. The quantization error of SOM and Batch-SOM is 8.97 and 9.24, respectively. K-means with a genetic algorithm were used to predict cluster levels in each building. By analyzing cluster levels, If-Then rules have been extracted, so needs that decision-makers determine the most energy-consuming buildings. In addition, this study helps decision-makers in the energy field to rationalize the consumption of occupants of public buildings in the times that consume the most energy and change energy suppliers to those buildings.

**Keywords**—Energy consumption in public buildings; self-organizing map; K-means; genetic algorithm; principal component analysis

## I. INTRODUCTION

The growing construction sector is struggling to cope with the increasing demand for energy despite efforts to develop sustainable buildings [1]. Therefore, improved energy efficiency and analysis of energy consumption patterns in buildings become necessary. This unveils the importance of understanding and classifying energy consumption patterns in buildings. For example, the more precise and pragmatic energy consumption profiles are computed, the better building energy quality evaluation becomes [2]. Energy consumption depends

on various factors such as building characteristics, energy prices, and climate conditions, amongst others [3]. Therefore, aiming to classify the energy consumption of buildings requires advanced computational intelligent approaches, particularly adopting the latest trends in machine learning, such as deep learning techniques, which exploit familiarity from historical data and can support decision-makers in the energy domain, creating a basis for styling new power allocation dispositions, particularly for public buildings areas [4].

Energy consumption in public buildings merits particular attention since it accounts for a large share of final energy consumption if we look at 2019 figures from the OECD (Organization for Economic Cooperation and Development) countries, which reached 27% in the European Union [5]. For example, public buildings consume nearly one-third of all electricity in Portugal, increasing by 35% from 1995 to 2019 [6]. Understanding this consumption means solving a complex problem involving physical, technological, and performance characteristics of the dwelling, the status of the demography, socio-economic factors, climate and weather conditions, and the behavior of the building's occupants [7]. Therefore, academic research in European countries, notably Portugal, needs help understanding the energy consumption patterns of public buildings.

In the past, we can find several data mining and machine learning techniques that have been used for energy consumption classification. Among those, clustering is considered one of the most applied techniques [8]. Clustering comprises splitting objects with similar styles into various groups [9]. Researchers have provided many manuscripts on classifying energy consumption into discrete levels. For instance, Gouveia [10] discovered electricity consumption profiles in households through clusters by combining smart meters and door-to-door surveys. His study used hierarchical clustering to divide household profiles and obtained three clusters. Hernandez et al. [11] presented a study to classify daily load curves in industrial parks by using a self-organizing map and k-means to determine the number of clusters. Ford and Siraj [12] presented a fuzzy c-means clustering to classify smart meter electricity consumption data to similar groups. Hodes et al. [13] presented a study to classify residential houses with similar hourly electricity using the k-means algorithm. Azaza [14] presented a method to find the most responsible energy consumers in the peak hour by using

hierarchical clustering and a self-organizing map. Al-Jarrah et al. [15] presented a method to discover power consumption in buildings using multi-layered clustering. K-mean has been utilized to partition power consumption profiles. Then, the authors discover different patterns of power consumption profiles. Furthermore, Cai et al. [16] presented a hybrid method to divide the electricity consumption of an entire region into various levels by using k-means with particle swarm optimization. To extract behavior in daily electricity consumption in households, Nordahl et al. [17] utilized the centroids of the generated clusters.

Most research focuses on the total energy consumption in different buildings by reviewing the analyzed literature. However, other factors that affect energy consumption, such as the consumption behavior of the occupants of these buildings at peak time or during empty hours (00h00-02h00; 06h00-08h00; 22h00-00h00), were neglected. In this paper, we follow this literature trend, and we propose an intelligent computing model capable of automatically classifying energy consumption into discrete levels, such as low, medium, and high. In this model, we can discover the different consumption patterns of public buildings across the country and visualize such patterns at different levels of the geographical organization and during the year, showing the different districts, municipalities, and parishes in which, the energy consumption is low, medium or high, in a certain period, helping to direct the occupant's behavior in such public buildings. The contribution of our paper has four dimensions:

1) Development of a novel hybrid model for classifying energy consumption in buildings (with an application to public buildings): the SOM, PCA, K-means (KM), and Genetic algorithm (GA), referred to as the SPKG model.

2) Evaluation of the performance and precision of the proposed model is trained and tested with real big data of energy consumption of public buildings in Portugal, collected in the years 2018 and 2019 (81 260 public buildings of 238 Portuguese cities).

3) Correlation coefficient analysis, to understand the relationship between the factors influencing energy consumption in buildings and determine the optimal factors amongst them.

4) A clustering and classification model of energy consumption levels in buildings, featuring a comparison between SOM and Batch-SOM based on PCA, in terms of quantization error, to select the optimal model between them and determine the optimal number of clusters in energy consumption in buildings. In our approach, we use the PCA algorithm to optimize SOM's weights, which helps to enhance the SOM model's fitting ability. Moreover, GA was used to find the optimal initial centroids in KM. This last technique predicts the cluster label in each building.

The paper is organized as follows. Section II presents our related work. In Section III, we present research questions and methodology: intelligent computing model. Section IV presents experimental results and discussion. Finally, in Section V, we conclude and suggest lines for further work.

## II. RELATED WORK

Putting public buildings that use the same amount of energy into similar groups is a key part of figuring out how much better or worse one building performs compared to similar buildings, like peers in the same group. Therefore, it is imperative to correctly identify these divisions to help the decision-maker in energy on three essential points: rationalizing the occupants of public buildings that consume much energy, determining the required amount of energy expected in the coming years, and changing energy providers in public buildings.

The most common methods for analyzing energy consumption in buildings are the different types of clustering methods [17]. Previous research analyzed raw meter data and used that data to represent energy consumption patterns using traditional statistical methods such as regression analysis and others [18,19]. The two most used clustering methods are K-means and Hierarchical clustering, which provide the most energy for occupancy and load forecasting [20 - 22]. Other machine learning methods are used to predict power consumption and loads, such as Artificial Neural Networks (ANN), Support Vector Machines (SVM), and K-Shape and other clustering methods [21 - 25].

Some important studies focused on finding an optimal way to understand occupancy schedules and user demand in different buildings, using anomaly detection and clustering methods [8, 26, 27, 28]. In anomaly detection, occupancy behavior is often used to design strategies that fit dynamic needs, user conditions, and interior space [9]. In addition, it helps design future buildings with a strategy that conserves wasted energy [29].

Other studies focus on measuring electricity consumption in buildings with their various activities using different methods of machine learning (i.e., decision trees [30] and stochastic frontier analysis [31]). These studies used intelligent methods to determine the different forms of electrical loads. In addition, it has been applied to more than 3000 residential and non-residential buildings.

Literature efforts are being conducted to find an intelligent computing model for clustering energy consumption in buildings using different factors that depend on the state of those buildings at different times and discovering the energy consumption patterns of occupants in such buildings [6]. Identifying and clustering the energy load patterns of occupants in public buildings based on such consumption profiles can be beneficial to stakeholders who aim to improve the energy efficiency of buildings effectively. K-means clustering is one of the methods used in the analyzed literature. However, it shows several issues. For example, K-means cannot group data where the groups are of varying volume and density [32]. Secondly, centroids can be pulled by outliers [33]. Finally, K-mean assumes that all variables have the same variance [34]. Consequently, our work tries to find a more accurate clustering method to overcome the limitations of the K-means clustering approach.

By analyzing previous works, we noticed the inability of these studies to find data that represents the occupants'

behavior of buildings at different times. Also, some papers use traditional statistical models like regression analysis and common clustering methods without paying attention to how well these models divide energy use into similar groups. The inaccurate classification of energy consumption leads to several ways to mislead the decision-maker: (1) the inability to find buildings that have high energy consumption; (2) the lack of anticipation of the energy required to cover the needs of public buildings adequately, and finally, the inability to identify the best energy providers. An energy consumption dataset was collected from Portuguese public buildings in 2018 and 2019 to remedy these shortcomings. This dataset was used to train and test a hybrid intelligent computing model to cluster energy consumption in public buildings. We believe that the decision-maker in the energy field can rely on this model to make sound decisions regarding energy consumption in public buildings and energy providers. In addition, there is a clear difference between the data used in this study and the data used in previous studies in terms of data quality and size, as the quality of detailed data on electricity consumption at various times of the day and the large size of data compared to previous studies. Moreover, in the preprocessing section, recent hybrid intelligent techniques such as isolation forest and interpolation methods were used that were not used in the same form and accuracy in previous studies. Furthermore, public buildings with high energy consumption have been determined in detail compared to previous works. Finally, recent hybrid intelligent techniques such as KM with GA were used to predict cluster labels. All these features make this study distinct from the rest of the previous studies.

### III. RESEARCH QUESTIONS AND METHODOLOGY

To properly frame our research, we raised the following research question:

- RQ1: What types of data sources and critical factors can be adopted to profile the energy consumption of buildings?
- RQ2: Which intelligent computing technique(s) can be adapted to identify the number of clusters in the given energy consumption dataset?
- RQ3: What general rules can be extracted to help the decision-maker rationalize energy consumption for public buildings?
- RQ4: What are the different and essential patterns discovered in the given energy consumption dataset?

To tackle the raised research question, we propose a hybrid approach (see Fig. 1), with a mixture of machine learning and optimization techniques, namely, (SOM [9]), (PCA [4]), KM, and GA [4], referred to as the SPKG model, able to discover different energy consumption patterns in buildings, with a proof of concept of its application to public buildings in Portugal.

In this section, we describe in detail our proposed model, which is composed of four main phases, as depicted in Fig. 1, namely:

- **Data Collection:** Our collected data includes energy consumption and building characteristics, such as (but not limited to): unique energy point of delivery ID, address of such a point of delivery, contracted electrical power, electricity consumption, and billing data with the month of consumption. The objective of this phase is to ensure that the units of measurement are consistent, that the sampling rates are adequate, that the time series is the same and synchronized over time, and that there were no structural changes during the data collection period.
- **Data Analysis and Pre-Processing:** In this phase, we analyze the data in detail and, if needed, transform it to expose its information content better. We adopt different mathematical techniques, namely, outlier removal with Isolation Forest (ISF) [35] and polynomial interpolation [31].
- **Feature Engineering:** In this phase, we find the optimal variables used to discover energy consumption patterns in (public) buildings, adopting a coefficients analysis approach [35].
- **Clustering Analysis:** In this phase, we fine-tune, apply, and evaluate our SPKG hybrid machine learning model, which can find clusters (each cluster corresponds to an energy consumption profile of buildings), and cluster the energy consumption profile in a particular building. Intelligent computing techniques, such as SOM and KM, are assessed and compared for automatic cluster discovery and the definition and classification of energy consumption behavior in (public) buildings.
- **Clustering Results:** In this phase, we tried to find three important results: generate energy consumption rules, determine the final number of clusters, and determine municipalities and Portuguese building activities that consume high, medium, and low energy consumption.

#### A. Data Collection

The data used in this study consists of the energy consumed in public buildings in Portugal, with the following characteristics: monthly data collected during the years of 2018 and 2019 in 77 996 buildings of various public sectors and 238 cities, reaching 2 775 082 records. After removing the records related to public lighting (since it is outside the scope of our study) and removing buildings that do not contain consumption data for the full observed period of 24 months, the number of records used in this study reached 1 222 695, corresponding to 26 624 public buildings.



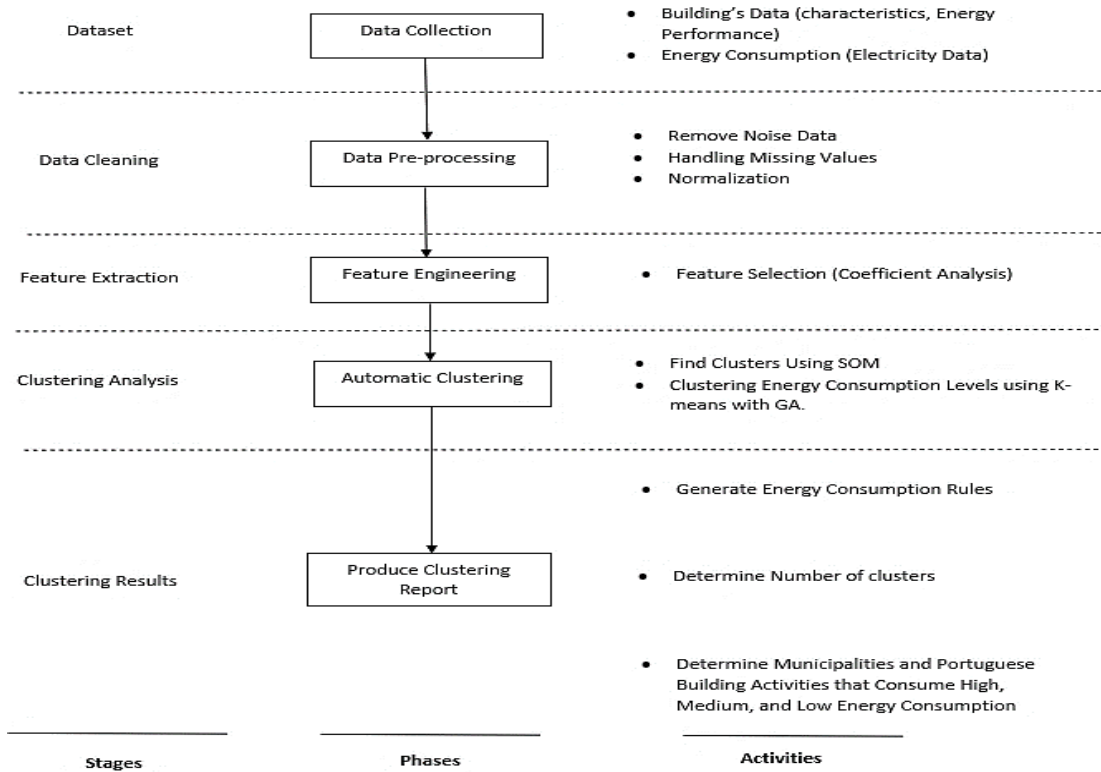


Fig. 1. Our proposed SPKG model for discovering energy consumption in public buildings

As mentioned, the dataset used in this study consists of two parts: the building characteristics and the actual energy consumption in these buildings (see Table I). Building characteristics include several attributes, namely:

- Unique energy is the point of the delivery ID of each building.
- Details of each building.

Energy consumption data, in our data set, includes:

- Actual active energy consumption in public buildings.
- Super empty: Active energy in the period 02h00-06h00 AM.
- Empty: Active Energy in the periods 00h00-02h00, 06h00-08h00, and 22h00-00h00.
- Outside empty: Lighting and plug loads that cannot be turned off.
- Peak: Active Energy in the periods 09h00-10h30 and 18h00-20h30.
- Full: Active Energy in the periods 08h00-09h00, 10h30-18h00, and 20h30-22h00.
- Total energy consumption: Active and Reactive Energy, where reactive energy is electrical energy that is stocked rather than transformed to some other form of energy and thus not "used" or "consumed."

TABLE I. DATASET DIMENSIONS OF ENERGY CONSUMPTION IN PUBLIC BUILDINGS

Dataset Dimensions	Attribute Name	Description
Characteristics of buildings	Unique Energy Point Delivery ID	The ID of each public building
	Business Partner	Identification of the institution that owns or rents the building.
	Building Address	Address of each building
	Municipality	City Location of each building
	Installation Type	Details of the electrical installation of each building
	Contracted Power	Power in MW has been agreed upon with the operator for each building.
	Year/Month	Consumption date
Energy consumption (Active Energy (KWh))	Simple	Total of active energy
	Super Empty	Active Energy (02h00-06h00)
	Empty	Active Energy (00h00-02h00; 06h00-08h00; 22h00-00h00)
	Outside Empty	Lighting and plug loads that cannot be turned off
	Peak	Active Energy (09h00-10h30; 18h00-20h30)
	Full	Active Energy (08h00-09h00; 10h30-18h00; 20h30-22h00)
	Total	Total of energy consumption (Active plus Reactive Energy)

## B. Data Preprocessing

Outlier detection and missing value imputation are the two primary processes in the data preprocessing for missing data utilizing the isolation forest and interpolation method. Here is a general description of the procedure [36 - 42]:

### Step 1: Outlier Detection using Isolation Forest

- Determine which features (columns) are missing data.
- For each characteristic, distinguish between the entire data (rows without missing values) and the partial data (rows with missing values).
- Using the whole data for each feature, isolate outliers using the isolation forest algorithm. A well-liked approach for anomaly identification called isolation forest isolates outliers by building random forests and calculating the typical number of splits required to isolate a data point.
- Establish a threshold to help you spot outliers. This may depend on the number of splits or a predetermined cutoff point.

### Step 2: Missing Value Imputation

- Use interpolation techniques to impute the missing values for the features that have missing data. Interpolation is a method that calculates the missing values from the data points already there.
- There are numerous interpolation techniques, including linear interpolation, polynomial interpolation, and methods tailored to time series, including forward-fill and backward-fill.

### Step 3: Combine Outlier Detection and Imputation

- We chose to keep outliers in the data after identifying them with the isolation forest.
- Use the selected polynomial interpolation technique to fill in the data gaps for the missing values.

## C. Feature Selection

This section aims to find the critical variables or factors in our energy consumption dataset. To overcome this problem, we used the T-test correlation coefficient. This statistical technique is used in literature to detect if two factors/variables are significant [43]. It can be helpful in our study. In our dataset, looking at pairwise correlations between the various variables (or factors) may propose a causal relation between two factors that we can investigate further. Eq. (1) computes the T-test value by assuming no correlation with  $\rho = 0$ , where, P refers to that; there is no relationship between variables [44].

$$t = r \sqrt{\frac{n-2}{1-r^2}} \quad (1)$$

In (1),  $n$  refers to the instances, and  $r$  represents the correlation coefficient of the energy consumption dataset. The importance of relevance is expressed in probability levels:  $p$  (e.g., significant at  $p = 0.05$ ). The degree of freedom for entering the t-distribution is  $n - 2$ . If the  $t$  value is less than the

critical value (CV) at a 0.05 significant level, the factor is not essential and is avoided [44].

In Algorithm 1, we build the correlation coefficients using the training dataset. In Steps 1 to 4, we calculate the correlation coefficients between the proposed factors. Step 6 to step 7 computes significant values by using the T-test. Finally, step 8 to step 10 finds the final list of energy consumption factors.

---

### Algorithm 1: Feature Selection Algorithm

---

Input:  $S(F_1, F_2, \dots, F_k, F_c)$  // a training data set

Output:  $S_{best}$  // the selected feature set

```
1. begin.
2. For I to k do
3. r = compute correlation coefficient ( $F_i, F_c$ )
4. End
   // let P = 0.05 significant level
   let P = 0 // assuming there is no significant correlation
5. For I to k do
6. t = compute significant values (r,p) for  $F_i$  // Eq.4
7. If t > CV // critical value
8.  $S_{list} = CV$ 
9.  $S_{best} = S_{list}$ 
10. End
11. End
12. Return  $S_{best}$ 
```

---

## D. Finding the Number of Clusters

To determine the optimal number of clusters in energy consumption data, we used three literature methods: Self-Organizing Map (SOM), the Elbow method, and the Bouldin & Davis method [14, 15]. These methods have been used in prior studies to find the optimal number of clusters, notably in energy consumption in buildings.

1) *Self-Organizing Map*: SOM is a specific class of neural networks utilized broadly as a clustering and visualization instrument in exploratory information analysis [45]. The main objective of SOM is to convert a complex high-dimensional discrete input space into a less low-dimensional discrete yield space by keeping the topology within the information but not the real separations [46, 47]. An unsupervised learning calculation employs a basic heuristic strategy for finding covered-up non-linear structures in high dimensional information [46].

The SOM method is adopted because it deals with big data accurately and effectively [45]. Contrary to the other methods mentioned, it better deals with small and medium-sized data [46]. Therefore, the SOM method determined the number of clusters in the energy consumption dataset. SOM is composed of three main processes: competition, cooperation, and adaptation [45].

The SOM network is composed of two layers, the input, and the output layer, as shown in Fig. 2. Each input variable is shown using an m-dimensional input vector [48]. In the output layer, the number of nodes indicates the most extreme number of clusters and impacts the precision and generalization capability of the SOM [45]. The arrangement of the SOM

begins with the initialization of the weight vectors [46]. Then, weights are joined that interface the input nodes to the output nodes and are overhauled through learning. Finally, to discover the best match unit (BMU), the spaces between an input ( $x$ ) and the weight vectors ( $w_i$ ) of the SOM are calculated by using various measurement methods, such as [47,49]:

- Manhattan distance.
- Chebyshev distance.
- Euclidean distance.
- Mahala Nobis distance
- Vector product, among other methods.

Euclidean distance is an approved measure in most scientific papers [47], as shown in Eq. (2):

$$d_i(t) = \|x(t) - w_i(t)\| \quad (2)$$

At the finish of the propinquity matching method (Determine the similarity between points in the dataset), the most excellent matching unit  $c$  at repetition  $t$  is identified by the minimum distance [45].

$$c(t) = \arg \min_i \|d_i(t)\| \quad (3)$$

By analyzing the weight vector  $w_i(t)$  of the winning neuron,  $i$  at iteration  $t$ , the overhauled weight vector  $w_i(t+1)$  at iteration  $(t+1)$  is determined by Using a discrete-time formalism in Eq. (4) [47].

$$w_i(t+1) = w_i(t) + \alpha(t) [x(t) - w_i(t)] \quad (4)$$

The weights ( $\alpha$ ) adjustment rate diminishes away from the winning node regarding the Spatio-temporal decay function [46].

$$h_{ci}(t) = \exp(-(d^*d)^{ci} / 2\sigma^2(t)) \quad (5)$$

where,

- $d$  is the lateral distance between the winning neuron  $c$  and the excited neuron  $i$ .
- $\sigma$  is the effective width or radius of the neighborhood at iteration  $i$ .

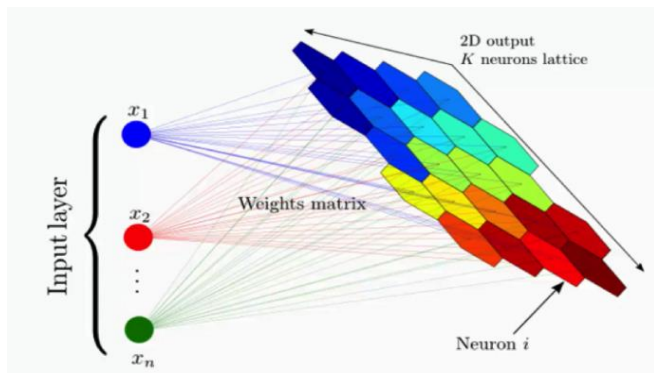


Fig. 2. Structure of SOM [45]

In Algorithm 2, we implemented a SOM network based on energy consumption data to determine the optimal number of clusters. Firstly, we have identified the lattice space of the  $10 \times$

$10$ , set weights based on random weights and PCA weights, set iterations from 100 to 1000. Secondly, pick the random points in energy consumption data, then find the best match point based on Eq. (5), set learning rate = 0.5, set neighborhood function = triangle, compute neighborhood distance weight matrix and modify SOM weight matrix, and finally, repeat from the step of picking random point ( $z$ ) until the maximum number of iterations is reached.

---

Algorithm 2: Main Idea of the SOM Network Training

---

**Input:** ECD  $\leftarrow$  the energy consumption data.

**Output:** USOM  $\leftarrow$  U-matrix of SOM network.

1.  $\beta \leftarrow$  initialize lattice nodes.
2.  $\Omega \leftarrow$  initialize weight vectors.
3.  $N \leftarrow$  Iteration count.
4. **For**  $i \leftarrow 1$  to  $N$  do
5.  $z \leftarrow$  picks a random point in ECD.
6.  $c \leftarrow \beta$  closest to  $z$ .
7. move the weight vector of  $c$  closer to  $z$ .
8. move the weight vectors of the neighbors of  $c$  slightly closer to  $z$ .
9. **End**

**Return** USOM

---

2) *Elbow method:* We can plot the curve indicating the average inner per cluster sum of squared error (SSE) distance vs the number of clusters to discover a visual "elbow", the ideal number of clusters. The average inner whole of squares is the average distance between focuses interior of a cluster [11], as shown in Eq. (6).

$$k = \sum_{r=1}^k \left( \frac{1}{n_r} + D_r \right) \quad (6)$$

Where:

- $k$  is the number of clusters,
- $n_r$  is the number of points in cluster  $r$ .
- $D_r$  is the sum of distances between all points in a cluster.

3) *Bouldin and davis method:* In Davis and Bouldin (DB), the score is characterized as the average similitude degree of each cluster with its most identical cluster. The similitude is the proportion of within-cluster separations to between-cluster separations. In this way, clusters that are more distant separated, and less scattered will result in a distant better score. The least score is zero, with lower values indicating superior clustering [13], as shown in Eq. (7) and (8) [16].

$$DB(c) = \frac{1}{k} \sum_{i=1}^k (\max_{j \leq k, j \neq i} D_{ij}), \quad k = |c| \quad (7)$$

$D_{ij}$  is the "within-to-between cluster distance ratio" for the  $i$ th and  $j$ th clusters.

$$D_{ij} = \frac{d_i^- + d_j^-}{d_{ij}} \quad (8)$$

where,  $d_i^-$  is the average distance between every data point in cluster  $i$  and its centroid, similar for  $d_j^-$ .  $d_{ij}$  is the Euclidean distance between the centroids of the two clusters.

E. K-Means with GA

GA is a research process inspired by Charles Darwin's theory of naturalist evolution. It is a process to select the fittest individuals to reproduce to create offspring of the next generation. GA is good at dealing with multiple points and is good in noisy environments; therefore, it quickly helps implement any fitness function such as Euclidean distance in the energy consumption dataset. GA was used to find the optimal centroids in KM to speed up convergence between energy consumption points through three fitness functions which are Euclidean distance (ED), Manhattan distance (MD), and Cosine distance (CD), as shown in formulas (2, 9, 10) [47]. Moreover, it helps to improve the accuracy of KM in our study.

MD indicates the sum of the absolute values of the differences of the coordinates. For example, if  $X = (E, M)$  and  $Y = (B, K)$ , the MD between  $X$  and  $Y$  is:

$$MD = |E - B| + |M - K| \tag{9}$$

CD calculates the cosine of the angle between vectors  $X$  and  $Y$  as shown below:

$$CD = \frac{X \cdot Y}{\|X\| \|Y\|} \tag{10}$$

Where:

- $\|X\|$  = Euclidean norm of vector,  $X = (X_1, X_2, \dots, X_n)$ .
- $\|Y\|$  = Euclidean norm of a vector,  $Y = (Y_1, Y_2, \dots, Y_n)$ .

KM aims to group identical data points as one cluster and detect underlying patterns. It has many challenges. First, determine the optimal number of previously determined clusters utilizing SOM. Second, determine the optimal centroid placement in each cluster utilizing GA. Thus, KM has been used to predict cluster labels in each building in ECD. In Algorithm 3, we constructed the improved KM using SOM and GA as inputs. From step 1 to step 4, improved KM tries to find the new centroid positions in each cluster for enhancing the accuracy of predicting the cluster label in each building in ECD.

Algorithm 3: Improved KM to predict cluster label in each building

---

```

Input: K = 3, // Specify the number of clusters using SOM
Initialize  $\sigma$  of centroids using GA.
Output:  $\beta \leftarrow$  predicting cluster label in each building in ECD
1. Repeat
2. Assign each point to its closest centroid.
3. Compute the new centroid of each cluster.
4. Until the centroid positions do not change.
Return  $\beta$ 
    
```

---

IV. EXPERIMENTAL RESULTS AND DISCUSSION

This section comprises four sections: data pre-processing, feature selection, finding the number of clusters, and finally, k-means with GA to produce energy consumption rules. We have used Python programming and the Scikit-learn library to implement the proposed algorithms.

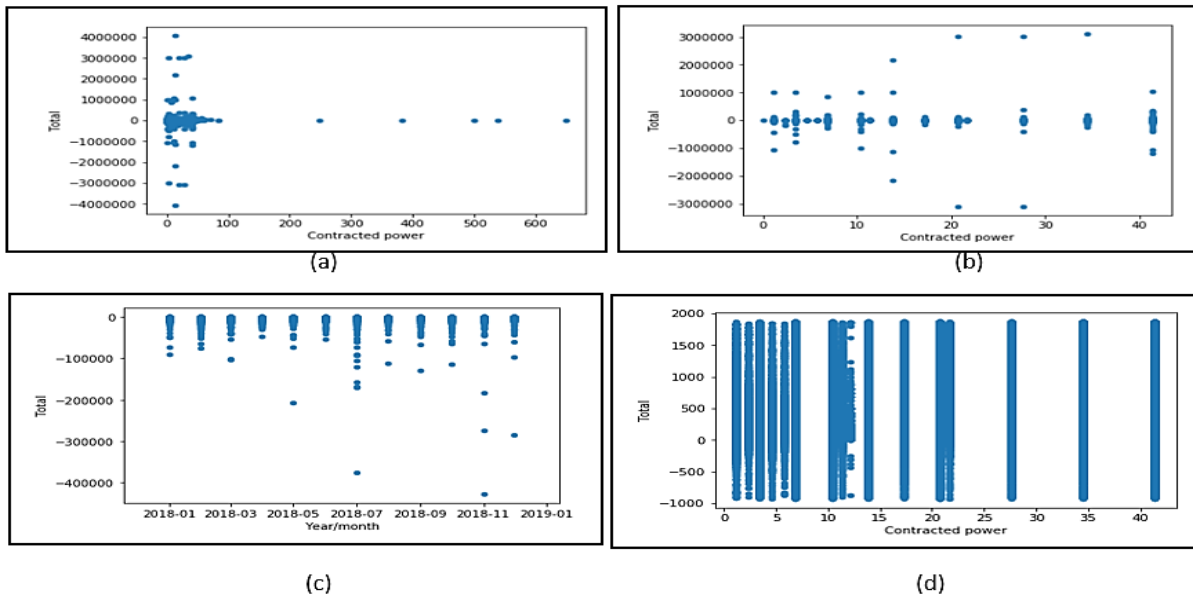


Fig. 3. Sample of data preprocessing.

A. Results of Data Preprocessing

Intelligent machine learning techniques always depend on the quality and efficiency of the dataset proposed in the study. Therefore, if the dataset provided is high quality and accurate, which helps to build and train an intelligent model with high efficiency. Furthermore, the energy consumption data is collected from a real-world environment. Therefore, it is

unstructured and incomplete. Thus, we always need the pre-processing data stage to remove noise and outliers. Data pre-processing has two stages. Fig. 3 shows the steps for pre-processing the energy consumption dataset in the first stage. Initially, (a) the sample of the raw dataset was displayed in terms of contracted power ( $X_i$ ) and total energy consumption ( $Y_i$ ); secondly, (b) Public buildings have been removed that

have several months less or more than 24 months, and public lighting buildings also have been removed because it is outside the scope of the study. Thirdly, (c) there are still public buildings that contain harmful and zero values. Fourthly, (d) outlier values have been removed using ISF, but harmful and zero values have also been removed.

After pre-processing, the final dataset was reached, which was relied upon to find the different patterns in energy consumption in public buildings. Fig. 4 shows the sample of the final data set between contracted power and total energy consumption.

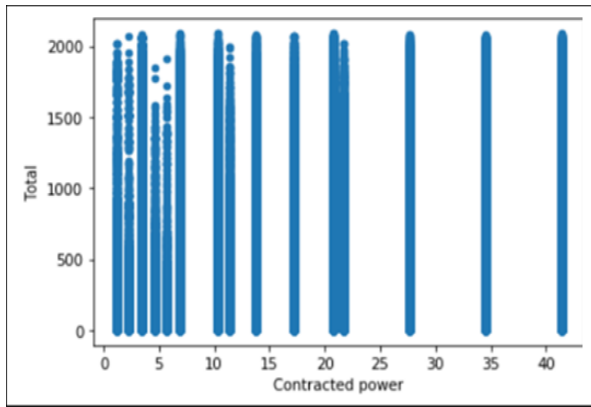


Fig. 4. Sample of Final Dataset

### B. Results of Feature Selection

The aim of this section is to show the results of the T-test correlation coefficient and find the critical factors in the energy

consumption dataset. Fig. 5 shows the relationships between energy consumption factors. We observed a relationship between contracted power with Full, Peak, Empty, outside empty, and total consumption, and there is also a relationship between Full and Peak. Moreover, there is a relationship between Empty and Outside Empty. Moreover, we can avoid the Super Empty factor because it contains null values in all the columns, and there is no relationship between it and all the other factors. Finally, there is a negative relationship between the Simple factor with Full, Peak, and Empty consumption.

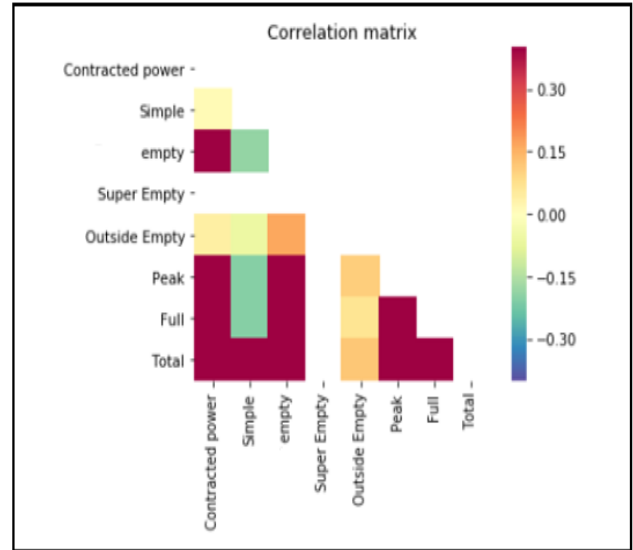


Fig. 5. The applied correlation coefficient in the energy consumption dataset.

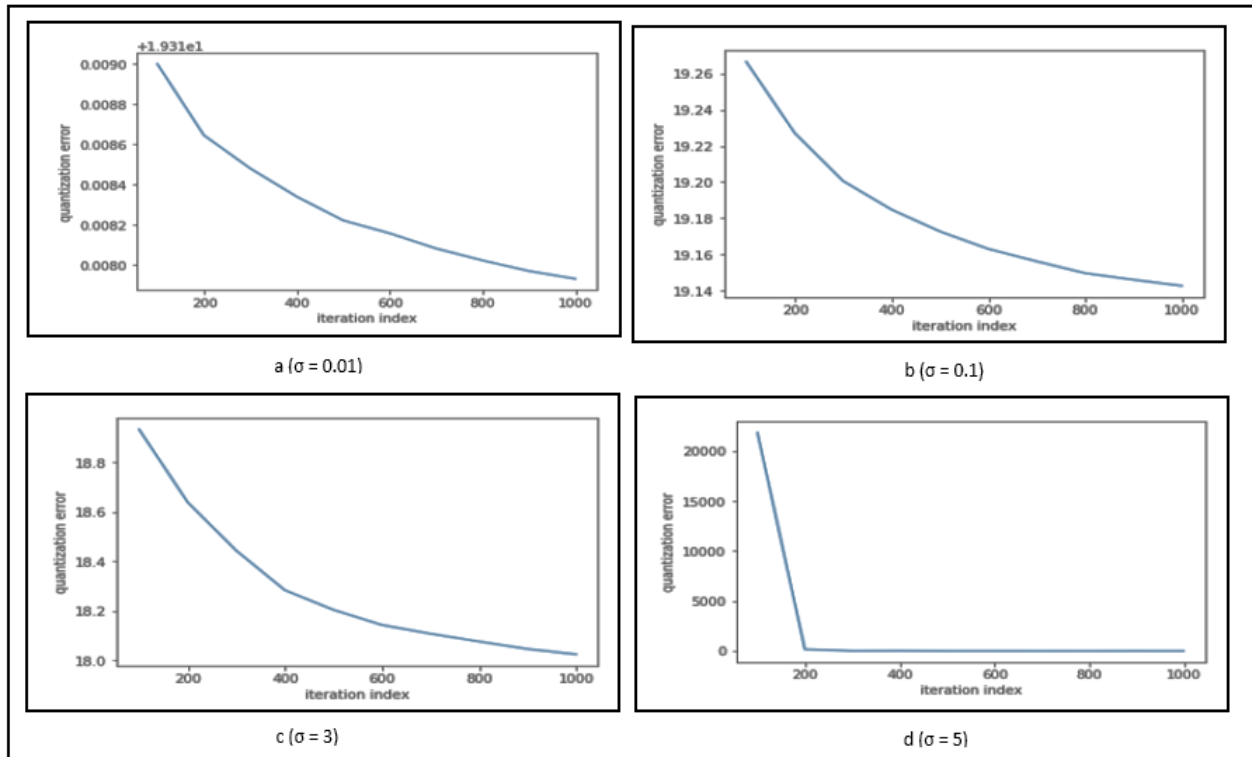


Fig. 6. q- error in random weights. Set iteration = 1000, (a)  $\sigma = 0.01$ , (b)  $\sigma = 0.1$ , (c)  $\sigma = 3$  and (d)  $\sigma = 5$ .

C. Results of Finding Number of Clusters

This section shows the results of three methods to find an optimal number of clusters: self-organizing map, Elbow method, and Bouldin and Davis method. A comparison was made on the weights of the SOM network in two different ways, the first utilizing random weights and the second through PCA weights. We set the iterations = 1000 and, we set sigma = 0.01, 0.1, 3 and 5. By comparing random weights and PCA weights, PCA weights are better than random weights in terms of quantization error (q- error), especially in iteration = 1000 and sigma = 3, as shown in Table II and Fig. 6 and 7.

The q- error expresses the squared distance (usually the average Euclidean distance) between input data x and their corresponding so-called BMU. Thus, the QE reflects the average distance between each data vector (X) and its BMU, as shown in Eq. (11):

$$q - error = 1/N \sum_{i=1}^N \|X_i - (BMU_{(i)})\| \quad (11) \quad [47]$$

The q- error appeared within Table II and Fig. 6 and 7 are midpoints for all data patterns. A comparative assessment of how this quantization is changed permits us to recognize distinctive clusters, which is one of the primary purposes of utilizing these techniques.

The SOM network was trained in two different ways based on PCA weights: random training SOM (RTSOM) and batch SOM (BSOM). The batch overhaul does not require a learning rate function. Typically, profitable since it reduces the number of required parameters. PCA weights with RTSOM (PCAW-RTSOM) are better than PCA weights with BSOM (PCAW-

BSOM) in terms of q- error. Q- error in PCAW-RTSOM and PCAW-BSOM is 8.97 and 9.24, respectively, as shown in Table III and Fig. 8.

TABLE II. A COMPARISON BETWEEN RANDOM WEIGHTS AND PCA WEIGHTS

SOM random weights			SOM PCA weights		
Iteration	Sigma	q- error	Iteration	Sigma	q- error
1000	0.01	19.32	1000	0.01	0.01
	0.1	19.14	0.1		216.85
	3	18.02	<b>3</b>		<b>13.14</b>
	5	20.13	5		16.39

TABLE III. A COMPARISON BETWEEN PCAW-RTSOM AND PCAW-BSOM

Iteration	PCAW-RTSOM	PCAW-BSOM
	q- error	
100	15.26	13.86
200	14.65	13.92
300	10.69	14.14
400	10.00	12.72
500	9.50	14.91
600	10.22	10.05
700	9.77	11.81
800	9.47	11.22
900	9.32	12.31
<b>1000</b>	<b>8.97</b>	<b>9.24</b>

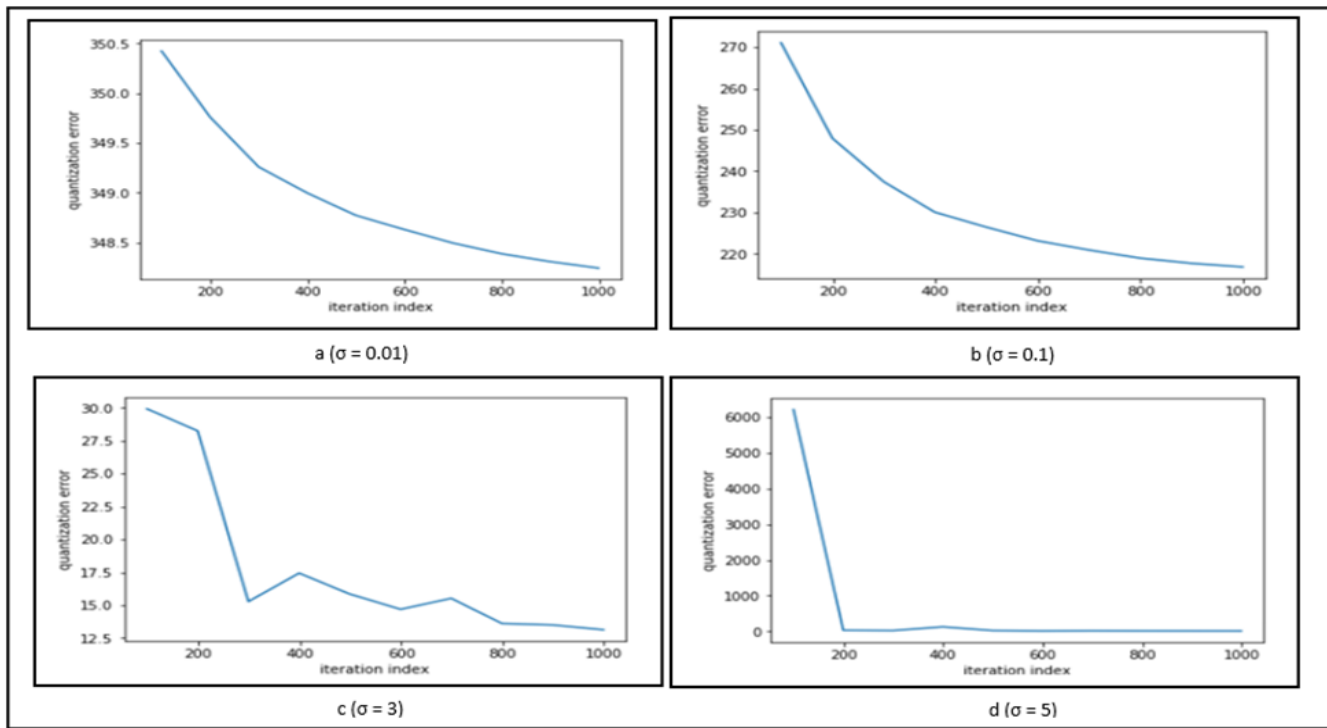


Fig. 7. q- error in PCA weights. Set iteration = 1000, (a)  $\sigma = 0.01$ , (b)  $\sigma = 0.1$ , (c)  $\sigma = 3$  and (d)  $\sigma = 5$ .

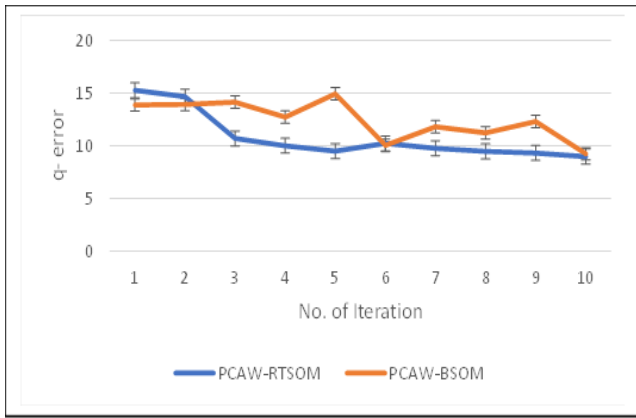
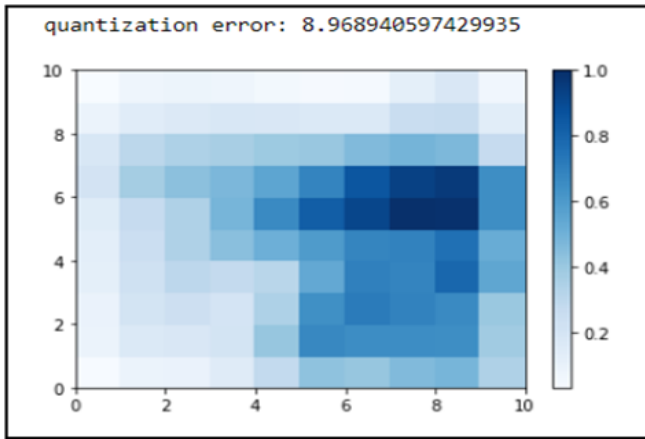


Fig. 8. A Comparison between PCAW-RTSOM and PCAW-BSOM.

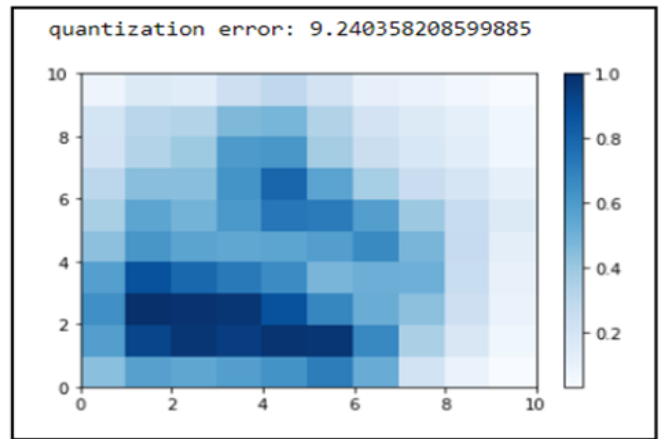
Fig. 9 shows the visualization of U-matrix in PCAW-RTSOM and PCAW-BSOM. In that U-matrix, we may determine three light color areas (white color) that match the minimum values in the U-matrix and indicate three clusters in the energy consumption dataset. These areas are detached by dark blue, which matches the segregation between the clusters.

We have obtained three clusters by implementing Elbow and Bouldin & Davis method in our energy consumption dataset, as shown in Fig. 10 and 11.

The U-matrix in SOM shows the distances between the points (points represent the energy consumption dataset) on the SOM. The dark areas in that U-matrix show the areas of the map where the points are far away from each other so, which represents the segregation between the clusters, and the lighter areas show fewer distances between the points so, which means the number of clusters. The Elbow method is computed as the intermediate of the squared distances from the cluster centers of the clusters. Typically, the Euclidean formula is utilized. In Bouldin & Davis method, to obtain the intra-cluster scuttle, we compute the average distance between each vector within the cluster and its centroid, which computes the Euclidean method between the centroid of the cluster. Finally, we could determine three clusters (low, medium, high consumption) in the energy consumption dataset by analyzing the SOM network, Elbow method, and Davis-Bouldin method.



(a)



(b)

Fig. 9. A Comparison between PCAW-RTSOM and PCAW-BSOM

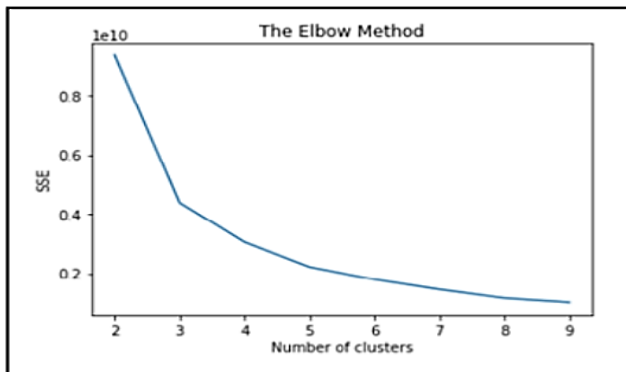


Fig. 10. Apply Elbow Method in Our Dataset

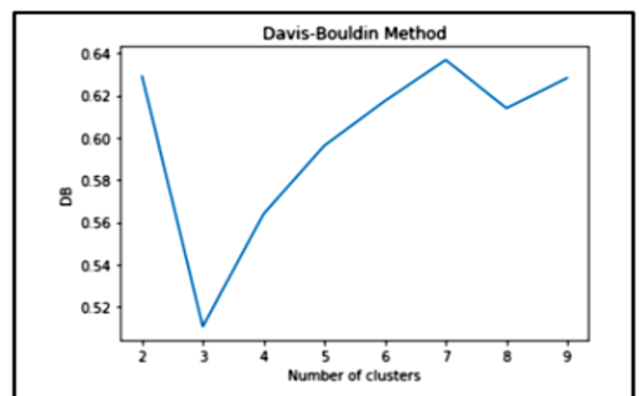


Fig. 11. Apply Davis-Bouldin Method in Our Dataset

D. K-Means with GA to Produce Energy Consumption Rules

In this section, we computed the distance between each cluster by two methods: The first method is K-means clustering with K-means++ initialization (KMCKI) and the second method is SPKG. GA has been implemented through the main parameters, as shown in Table IV. There are three methods to compute distances between clusters: ED, MD, and CD. We compared the performance between KMCKI and SPKG in terms of standard error (SE), as shown in formula 12, and standard deviation. CD with SPKG is better than all methods, as shown in Table V. Thus, this study relied on CD with SPKG to predict cluster labels in each building in ECD and detect underlying patterns.

$$SE = \frac{STDEV(\Omega)}{\sqrt{COUNT(\Omega)}} \quad (12) \quad [11]$$

Where:

STDEV = Standard deviation

$\Omega$  = Distances between each center of clusters.

It is an important step to visualize big data analytics. The clustering outputs have been shown in different methods to facilitate decision-makers and stakeholders in the energy field in Portugal to take suitable decisions in energy consumption in public buildings. In addition, ECD has the significant factor of contracted power, which is very useful in understanding how much energy is consumed during different times in the day in each public building. Fig. 12 shows a sample analysis of the various visualizations that show the dimensions used in the ECD through CD with SPKG.

By analyzing the clustering results, several essential rules have been extracted to assist stakeholders in the energy sector in Portugal in identifying the different styles of public buildings, as shown in Table VI. Energy consumption rules help the decision-maker identify public buildings that need guidance for their occupants and change the energy suppliers for those buildings.

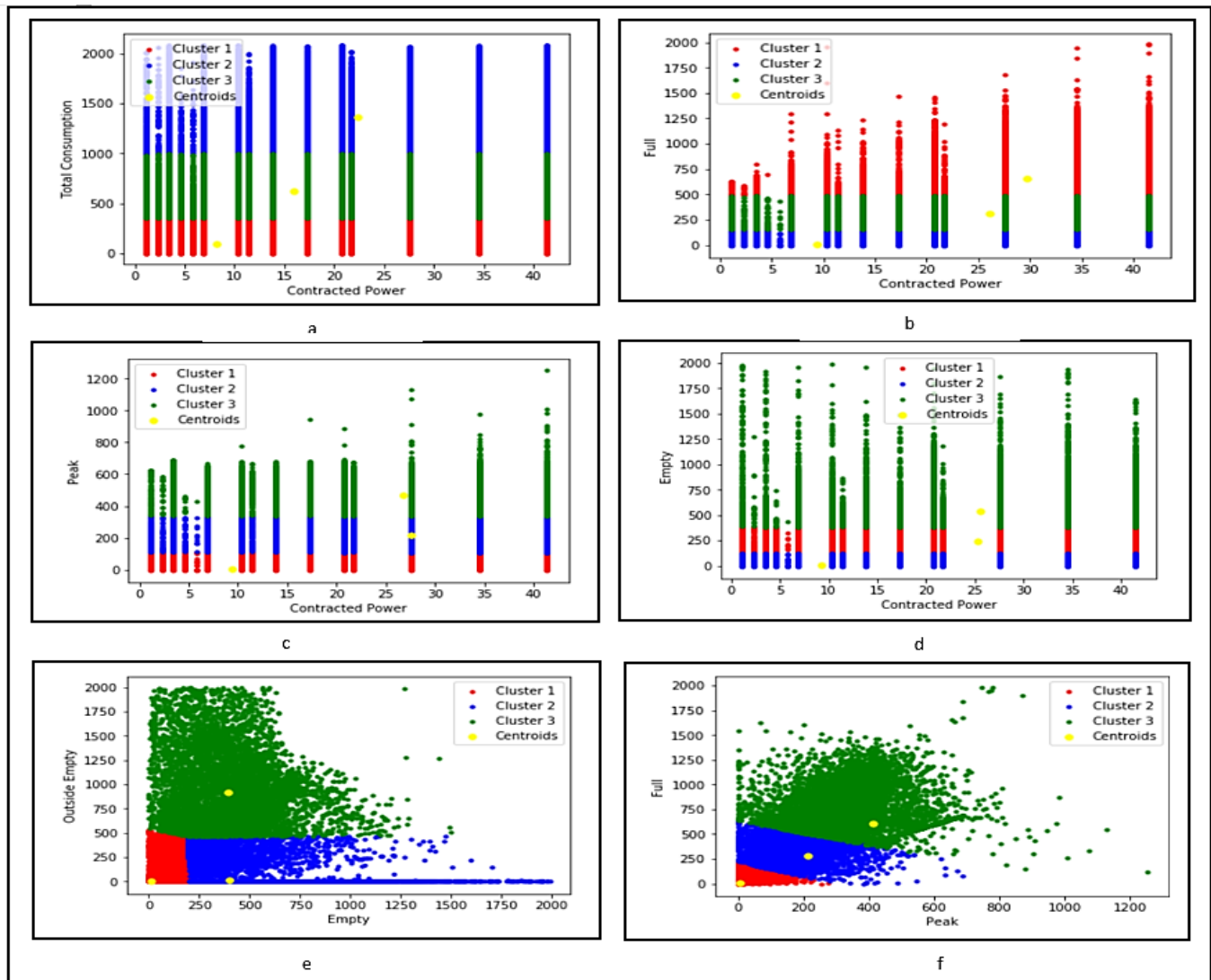


Fig. 12. Sample of clustering results.



TABLE IV. GA PARAMETERS

No	Parameters	Value
1	Population Size	ECD
2	Crossover Probability	0.5
3	Crossover type	Two points
4	Mutation Probability	0.6
5	Mutation type	Bit flip
6	Number of Iterations	100

TABLE V. A COMPARISON BETWEEN KMCKI AND SPKG IN TERMS OF SE AND STDEV

No	Method	SE	STDEV
1	ED with Kmeans++ (EDK)	93.19	465.99
2	MD with Kmeans++ (MDK)	184.14	920.73
3	CD with Kmeans++ (CDK)	0.004	0.021
4	ED with SPKG	88.49	442.45
5	MD with SPKG	174.94	874.71
6	CD with SPKG	<b>0.002</b>	<b>0.012</b>

TABLE VI. SAMPLE OF ENERGY CONSUMPTION RULES

No	Rules
1	Total<359 AND Full<157 Then cluster 1 (low energy consumption)
2	Total<359 AND Peak<111 Then cluster 1 (low energy consumption)
7	359<Total<992 AND 245<Outside empty<878 Then cluster 2 (medium energy consumption)
8	359<Total<992 AND 123<Empty<386 Then cluster 2 (medium energy consumption)
9	Total>=993 AND Full>=484 Then cluster 3 (high energy consumption)
10	Total>=993 AND Peak>=341 Then cluster 3 (high energy consumption)
14	157<Full<484 AND 111<Peak<341 Then cluster 2 (medium energy consumption)
15	Full>=484 AND Peak>=341 Then cluster 3 (high energy consumption)
16	Outside empty<245 AND Empty<123 Then cluster 1 (low energy consumption)
17	245< Outside empty <878 AND 123<Empty<386 Then cluster 2 (medium energy consumption)
18	Outside empty >=878 AND Empty>=386 Then cluster 3 (high energy consumption)
19	Total<359 AND Full<157 AND Peak<111 AND Outside empty<245 AND Empty<123 Then cluster 1 (low energy consumption)
21	Total>=993 AND Full>=484 AND Peak>=341 AND Outside empty>=878 AND Empty>=386 Then cluster 3 (high energy consumption)

Fig. 12 was better for detecting energy consumption levels; however, it could not determine the months in which energy consumption increases, as well as the months in which energy consumption decreases. Monthly consumption patterns show broader details of energy consumption by an occupant in public buildings. For ECD, the energy consumption levels were

determined based on cluster label predictions, as shown in Fig. 13. Fig. 13 shows a noticeable increase in electricity consumption in January, February, November, and December. In addition, energy consumption levels decreased in June and July. It also helps the decision-maker identify the months of increased energy consumption for public buildings. Thus, the occupants of these buildings are guided promptly.

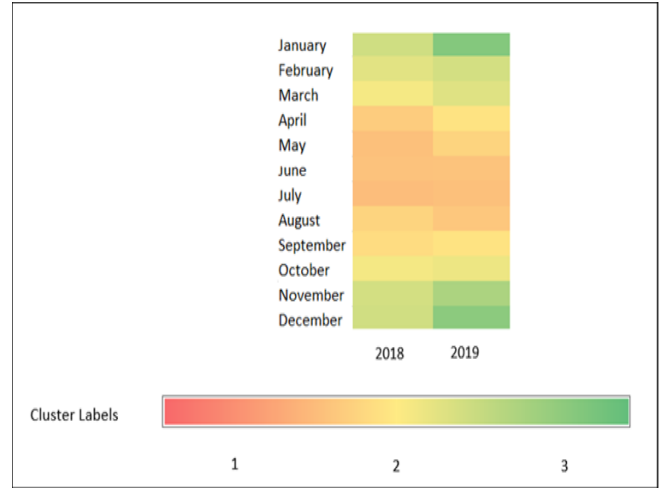


Fig. 13. Monthly energy consumption patterns captured in different clusters for the ECD.

Fig. 14 and Table VII show municipalities and Portuguese public buildings activities that contain the number of buildings that consume low energy at different times. Three municipalities contain public buildings that consume little energy in Fig. 14, such as 'LOULE', 'SANTA MARIA DA FEIRA', and 'LISBON'. In addition, Table VII shows Portuguese public buildings activities that consume little energy such as: 'INFRAESTRUTURAS PORTUGAL SA', 'GUARDA NACIONAL REPUBLICANA', and 'INSTITUTO SEGURANCA SOCIAL'.

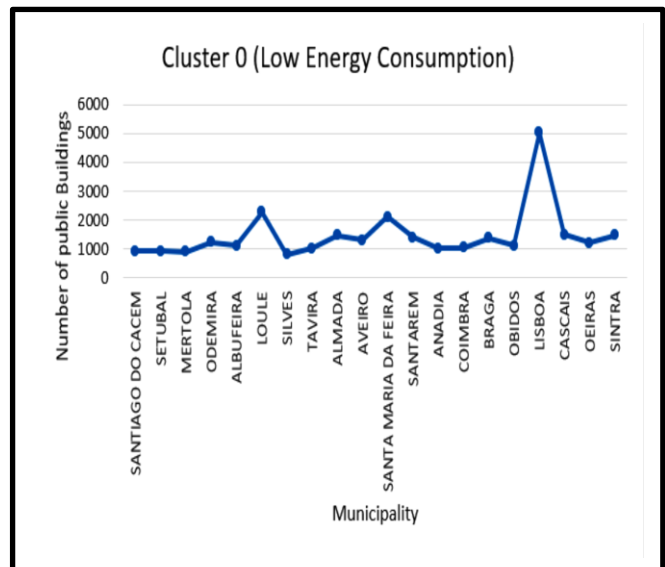


Fig. 14. Sample of Municipalities that Consume Low Energy Consumption

TABLE VII. SAMPLE OF PUBLIC BUILDINGS THAT CONSUME LOW ENERGY IN EACH MUNICIPALITY

Public buildings	Municipality	CACEM	SETUBAL	MERTOLA	ODEMIRA	ALBUFEIRA	LOULE	SILVES	TAVIRA	ALMADA	A VEIRO	MARIA FEIRA	SANTAREM	ANADIA	COIMBRA	BRAGA	OBIDOS	LISBOA	CASCAIS	OEIRAS	SINTRA
INFRAESTRUTURAS PORTUGAL SA		45	28	0	3	3	37	40	31	0	16	19	49	0	99	48	4	46	15	6	10
INSTITUTO SEGURANCA SOCIAL		3	0	21	7	6	12	12	0	13	0	23	0	0	0	0	0	0	0	62	5
ADMINISTRACAO REGIONAL SAUDE CENTRO IP		0	0	0	0	0	0	0	0	0	52	0	0	63	81	0	0	0	0	0	0
GUARDA NACIONAL REPUBLICANA		45	21	17	27	21	47	7	0	29	0	0	0	0	1	22	0	9	0	0	8

TABLE VIII. SAMPLE OF PUBLIC BUILDINGS THAT CONSUME MEDIUM ENERGY IN EACH MUNICIPALITY

Public building	Municipality	MONTEJO	CACEM	ALMODOV	MERTOLA	ODEMIRA	ALBUFEIRA	ALCOUTIM	CASTRO	LOULE	TAVIRA	MARIA FEIRA	MONTENMO	TORRES VEDRAS	BRAGANCA	VISEU	PORTO	VILA NOVA	LISBOA	CASCAIS	OEIRAS
IHRU INSTIT DA HABIT E REABILITACAO URBANA IP		0	105	0	0	0	0	0	0	23	0	0	0	0	0	0	2211	161	46	23	0
INFRAESTRUTURAS PORTUGAL SA		45	98	0	0	47	5	19	16	78	128	404	158	142	0	131	506	444	60	19	19
MUNICIPIO PORTO		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9797	0	0	0	0
MUNICIPIO OEIRAS		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8872

TABLE IX. SAMPLE OF PUBLIC BUILDINGS THAT CONSUME HIGH ENERGY IN EACH MUNICIPALITY

Public building	Municipality	SINTRA	ODEMIRA	ALBUFEIRA	LOULE	SILVES	ALMADA	A VEIRO	MARIA DA FEIRA	SANTA AZEIS	COIMBRA	SOUR	GUIMARAES	BRAGA	BARCELOS	MIRANDELA	LEIRIA	VILA NOVA DE GAIA	CASCAIS	LISBOA	OEIRAS
GUARDA NACIONAL REPUBLICANA		1	31	22	39	20	31	0	0	0	0	0	0	20	1	18	0	29	0	24	0
ADMINISTRACAO REGIONAL SAUDE CENTRO IP		0	0	0	0	0	0	70	0	0	115	54	0	0	0	0	106	0	0	0	0
ADMINISTRACAO REGIONAL SAUDE NORTE		0	0	0	0	0	0	0	272	42	0	0	50	102	53	0	0	165	0	0	0
AUTORIDADE TRIBUTARIA ADUANEIRA		11	7	0	19	0	14	0	28	0	0	0	3	0	2	18	0	0	13	23	0
INSTITUTO SEGURANCA SOCIAL		17	14	16	9	17	0	0	3	0	0	2	0	0	20	0	0	16	0	0	5

Fig. 15 and Table VIII show the municipalities and Portuguese public buildings activities that contain the number of public buildings that consume energy on average between low and high consumption at different times. In Fig. 15, three municipalities contain public buildings that consume energy reasonably, such as 'PORTO', 'LISBOA', and 'OEIRAS'. In addition, Table VIII shows Portuguese public buildings

activities that consume energy reasonably, such as: 'IHRU INSTIT DA HABIT E REABILITACAO URBANA IP', 'INFRAESTRUTURAS PORTUGAL SA', and 'MUNICIPIO PORTO'.

Fig. 16 and Table IX show the activities of municipalities and Portuguese public buildings containing the number of public buildings that consume high energy at different times. In

Fig. 16, four municipalities contain public buildings that consume high energy, such as: 'LOULE', 'SANTA MARIA DA FEIRA', 'VILA NOVA DE GAIA', and 'LISBOA'. In addition, Table IX shows Portuguese public buildings activities that consume high energy such as: 'GUARDA NACIONAL REPUBLICANA', 'ADMINISTRACAO REGIONAL SAUDE CENTRO IP', 'ADMINISTRACAO REGIONAL SAUDE NORTE', and 'AUTORIDADE TRIBUTARIA E ADUANEIRA'.

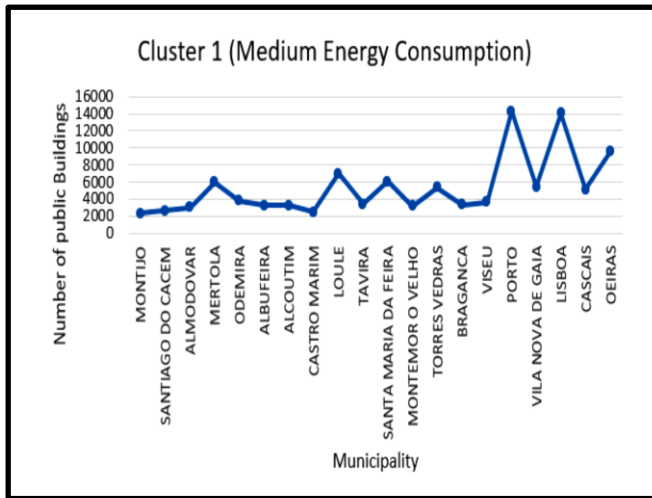


Fig. 15. Sample of Municipalities that Consume Medium Energy Consumption

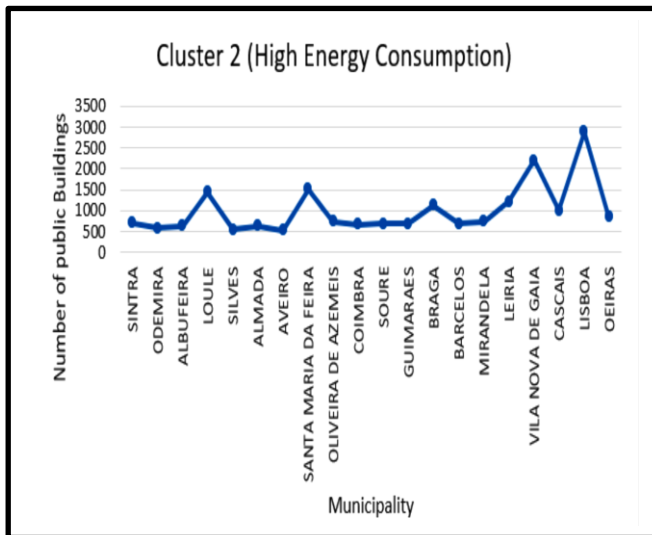


Fig. 16. Sample of Municipalities that Consume High Energy Consumption

This accurate analysis helps the decision-maker identify the municipalities and Portuguese public buildings activities that need to guide their consumers and change their energy providers.

By analyzing Fig. 14 to 16 and Tables VII to IX, municipalities such as 'LISBOA' and 'LOULE' contain public buildings with low, medium, and high energy consumption. In addition, there are Portuguese public buildings activities such

as 'INFRAESTRUTURAS PORTUGAL SA' that consume low and medium energy. Therefore, we seek to find the distribution of the number of public buildings with different activities with low, medium, and high energy consumption over the different municipalities.

Tables VII, VIII, and IX show a sample of the public buildings located within each municipality. Knowing that each building has more than one location appears 24 times, distributed over 24 months over two years, 2018 and 2019.

By analyzing Fig. 17, the number of public buildings in these Municipalities increased in certain months in 2018 and 2019 as follows:

- LOULE: Aug-18, Oct-18, Jan-19, Feb-19, Mar-19, and Oct-19.
- SANTA MARIA DA FEIRA: Feb-18, Mar-18, Apr-18, May-18, Jun-18, Nov-18, Jan-19, and Feb-19.
- BRAGA: May-18, Aug-18, Oct-18, Jan-19, Mar-19, Apr-19, and May-19.
- VILA NOVA DE GAIA: Aug-18, Sep-18, Oct-18, Nov-18, and Jan-19 to Oct-19.
- LISBOA: Feb-18 to Nov-18, and Jan-19 to Dec-19

Regarding answering our research questions, and starting from RQ1, which aimed to collect public buildings energy consumption data in Portugal, and to find which where the critical factors in such dataset that could helped us in profiling such consumption, we were able to obtain aggregated monthly data for the years 2018 and 2019, regarding 77 996 buildings of various public sectors in 238 cities in Portugal, reaching 2 775 082 records. We concluded that all factors (variables) of the collected data are critical for the mentioned profiling, except for the super empty variable. Our RQ2 aimed to find the more appropriate intelligent computing techniques, for the preparation of the energy consumption dataset to proceed with further clustering analysis. Answering to this question, we adopted different mathematical techniques to that aim, namely, outlier removal with Isolation Forest and polynomial interpolation. With a dataset ready for clustering analysis, we raised RQ3, seeking first, to identify the number of clusters in the given energy consumption dataset, where we adopted literature techniques such as Self Organizing Map (SOM), the Elbow method, and the Davis – Bouldin method, and then to propose a novel and optimized hybrid model for classifying (labelling) energy consumption in buildings. This model includes a mix of different techniques, namely, SOM, Principal Component Analysis (PCA), K-means (KM), and Genetic algorithm (GA), is referred to as the SPKG model, and was applied successfully to our dataset, predicting the cluster label (low, medium, or high consumption) of each building. With a set of labelled buildings at hand, we turned our attention to RQ4, targeting to discover essential patterns and general rules in such labelled dataset, that could help the decision-maker to rationalize energy consumption. Therefore, we analysed the clustering results and came up with a set of rules that can help the characterization of energy consumption of a given public building in Portugal.

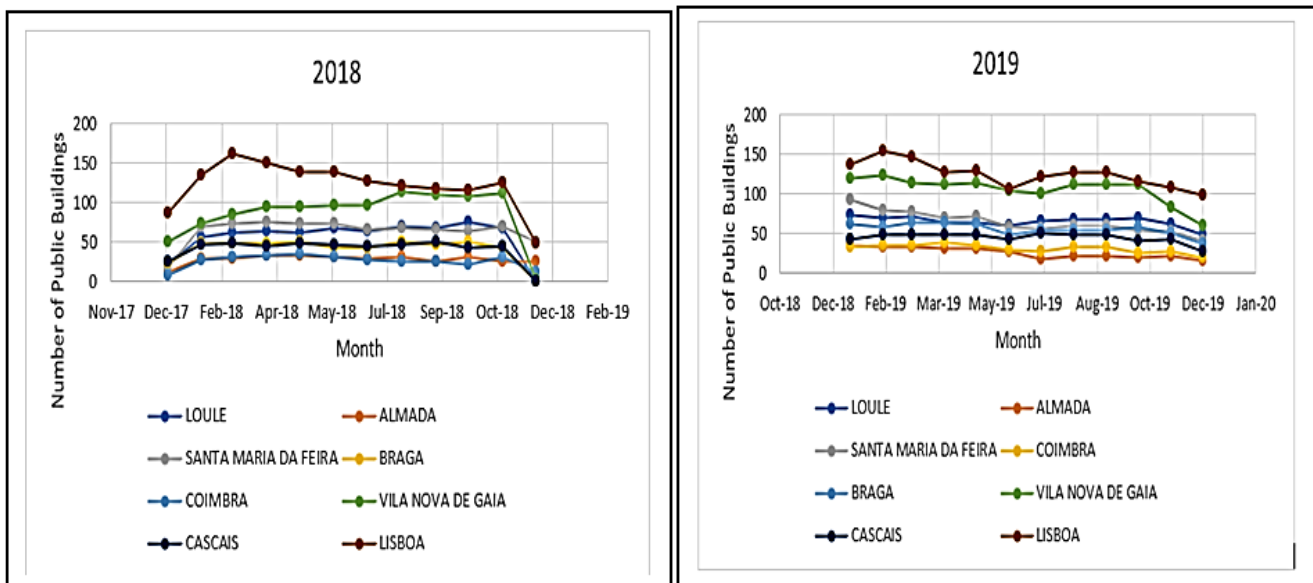


Fig. 17. Sample of Public Buildings in Different Municipalities that Consume High Energy in 2018 and 2019.

We have compared our results with state-of-the-art methods in the literature related to our work, in terms of using the K-means algorithm. M. Azaza study [14] and Al-Jarrah study [15]. The Standard Error (SE) of clustering in M. Azaza [14] and Al-Jarrah [15] is 28.3 and 22.5, respectively. However, SE in our study is 0.002. Therefore, our study outperforms state-of-the-art methods in previous work, in terms of SE of the K-means algorithm.

## V. CONCLUSION AND FUTURE WORK

This paper presented a novel hybrid intelligent model for classifying the energy consumption level (low, medium, high) of buildings that was tested in a dataset of energy consumption of Portuguese public buildings. To frame our research, we raised four research questions that were properly answered. To understand our data, a correlation coefficient analysis was used to find the critical factors (variables) that influence energy consumption of public buildings and understand the relationship between those factors. In a data preparation step, an isolation forest was used to remove outliers in the dataset. Additionally, an interpolation method was used to find compensation values or estimate unknown values using related known values. As for our modelling approach, aiming at labelling the energy consumption level of each building, we first computed the number of clusters of energy consumption in the dataset, and SOM, the Elbow method, and Davis-Bouldin method all agreed in 3 as the figure for the found number of clusters (corresponding to low, medium, high consumption).

Then we used K-means with a Genetic Algorithm to predict the energy consumption cluster level of each building. This study provides contributions in four aspects. The first one considers factors that influence the energy consumption of buildings. The second one provides a novel model for classifying energy consumption of public buildings into levels (e.g., low, medium, and high). The third one provides analysis on real big data of the energy consumption of public buildings in Portugal, in the years of 2018 and 2019 (77 996 public

buildings in 238 Portuguese cities). As an example, we were able to identify the municipalities that consume high energy levels. We have also identified monthly energy consumption patterns of buildings of the years of 2018 and 2019. The last aspect extracts proper scientific If-Then rules to help decision-makers rationalize the energy-consuming and determine the most energy-consuming public buildings, from a set of 3 values (low, medium, or high consumption).

Together, all these results may help the decision-maker to evaluate the public building's future energy requirements, and rationalize the occupants of those buildings, with the correct energy consumption behaviours.

As a recommendation for future work, we can think of using other techniques, such as statistical methods like multiple linear regression or logistic regression to find critical factors that influence energy consumption of public buildings. We could combine SOM and other optimization techniques (grey wolf, lion, and whale optimization), aiming find the optimal number of clusters of the energy consumption data of buildings. In addition, combining clustering and optimization techniques (grey wolf, lion, and whale optimization) could yield better prediction of cluster labels as for predicting the amount of energy consumption of buildings, this study follows the recent literature trend and suggests adopting machine learning approaches from the family of deep learning techniques, such as long short-term memory, convolutional neural networks, or deep forest.

## ACKNOWLEDGMENT

This work has been supported by Portuguese funds through FCT-Fundação para a Ciência e Tecnologia, Instituto Público (IP), under the project FCT UIDB/04466/2020 by Information Sciences and Technologies and Architecture Research Center (ISTAR-IUL), and this work has also been supported by Information Management Research Center (MagIC)-Information Management School of NOVA University Lisbon.

REFERENCES

- [1] T. A. Nguyen and M. Aiello, "Energy intelligent buildings based on user activity: a survey", *Energy and Buildings*, vol. 56, no. 1, pp. 244–257, 2013.
- [2] M. Zhang and C. Y. Bai, "Exploring the influencing factors and decoupling state of residential energy consumption in Shandong", *Journal of Cleaner Production*, vol. 194, no. 1, pp. 253–262, 2018.
- [3] N. Javaid, I. Ullah, M. Akbar, Z. Iqbal, F. Khan et al., "An intelligent load management system with renewable energy integration for smart homes", *IEEE Access*, vol. 5, no. 1, pp. 13587–13600, 2017.
- [4] K. Li, C. Hu, G. Liu and W. Xue, "Building's electricity consumption prediction using optimized artificial neural networks and principal component analysis", *Energy and Buildings*, Vol. 108, no. 4, pp. 106–113, 2015.
- [5] D. Zhao, M. Zhong, X. Zhang and X. Su, "Energy consumption predicting model of VRV (variable refrigerant volume) system in office buildings based on data mining", *Energy*, Vol. 102, no. 1, pp. 660–668, 2016.
- [6] E. Agência, "Energy efficiency trends and policies in Portugal", *Agência para a Energia*, Vol. 1, no. 1, pp. 234–251, 2018.
- [7] G. Shi, D. Liu and Q. Wei, "Energy consumption prediction of office buildings based on echo state networks", *Neurocomputing*, Vol. 126, no. 1, pp. 243–264, 2016.
- [8] S. Naji, A. Keivani, S. Shamshir, J. Alengaram, Z. Jumaat et al., "Estimating building energy consumption using extreme learning machine method", *Energy*, Vol. 97, no. 2, pp. 506–516, 2016.
- [9] J. Massana, C. Pous, L. Burgas, J. Melendez and J. Colomer, "Short-term load forecasting for non-residential buildings contrasting artificial occupancy attributes", *Energy and Buildings*, Vol. 130, no. 4, pp. 519–531, 2016.
- [10] J. P. Gouveia and J. Seixas, "Unravelling electricity consumption profiles in households through clusters: combining smart meters and door-to-door surveys", *Energy and Buildings*, Vol. 116, no. 2, pp. 666–676, 2016.
- [11] L. Hernández, C. Baladrón, J. Aguiar, B. Carro and A. Sánchez, "Classification and clustering of electricity demand patterns in industrial parks", *Energies*, vol. 5, no. 1, pp. 5215–5228, 2012.
- [12] V. Ford and A. Siraj, "Clustering of smart meter data for disaggregation", In *Proceedings of the 2013 IEEE Global Conference on Signal and Information Processing*, Austin, TX, USA, pp. 507–510, 2013.
- [13] D. Rhodes, J. Cole, R. Upshaw, F. Edgar, E. Webber et al., "Clustering analysis of residential electricity demand profiles", *Applied Energy*, vol. 135, no. 4, pp. 461–471, 2014.
- [14] M. Azaza and F. Wallin, "Smart meter data clustering using consumption indicators: responsibility factor and consumption variability", *Energy Procedia*, vol. 142, no. 4, pp. 2236–2242, 2017.
- [15] Y. Al-Jarrah, Y. Al-Hammadi, D. Yoo and S. Muhaidat, "Multi-layered clustering for power consumption profiling in smart grids", *IEEE Access*, Vol. 5, no. 1, pp. 18459–18468, 2017.
- [16] H. Cai, S. Shen, Q. Lin, X. Li and H. Xiao, "Predicting the energy consumption of residential buildings for regional electricity supply-side and demand-side management", *IEEE Access*, Vol. 7, no. 1, pp. 30386–30397, 2019.
- [17] C. Nordahl, V. Boeva, H. Grahn and P. Netz, "Profiling of household residents' electricity consumption behavior using clustering analysis", In *Proceedings of the International Conference on Computational Science*, Faro, Portugal, pp. 779–786, 2019.
- [18] R. Granell, C. J. Axon and D. C. Wallom, "Impacts of raw data temporal resolution using selected clustering methods on residential electricity load profiles", *IEEE Transactions on Power Systems*, Vol. 30, no. 1, pp. 3217–3224, 2015.
- [19] M. Christ, N. Braun, J. Neuffer and A. W. Kempa-Liehr, "Time series feature extraction on basis of scalable hypothesis tests (tsfresh – a python package)", *Neurocomputing*, Vol. 307, no. 4, pp. 72–77, 2018.
- [20] C. Miller, Z. Nagy and A. Schlueter, "A review of unsupervised statistical learning and visual analytics techniques applied to performance analysis of non-residential buildings", *Renewable and Sustainable Energy Reviews*, Vol. 81, no. 4, pp. 1365–1377, 2018.
- [21] D. Hsu, "Comparison of integrated clustering methods for accurate and stable prediction of building energy consumption data", *Applied Energy*, Vol. 160, no. 1, pp. 153–163, 2016.
- [22] A. Al-Wakeel, J. Wu and N. Jenkins, "K - means based load estimation of domestic smart meter measurements", *Applied Energy*, Vol. 194, no. 2, pp. 333–342, 2017.
- [23] A. S. Ahmad, M. Y. Hassan, M. P. Abdullah, H. A. Rahman, F. Hussin et al., "A review on applications of ANN and SVM for building electrical energy consumption forecasting", *Renewable and Sustainable Energy Reviews*, Vol. 33, no. 1, pp. 102–109, 2014.
- [24] D. Zhikuen, W. Zhan, T. Hu and H. Wang, "A comprehensive study on integrating clustering with regression for short-term forecasting of building energy consumption: case study of a green building", *Buildings*, Vol. 12, no. 10, pp. 1–20, 2022.
- [25] Z. Chen, F. Xiao, F. Guo, F. Zhang, J. Yan et al., "Interpretable machine learning for building energy management: a state-of-the-art review", *Advances in Applied Energy*, Vol. 9, no. 1, pp. 1–19, 2023.
- [26] T. Zhao, C. Zhang, T. Ujeed and L. Ma, "Methods on reflecting electricity consumption change characteristics and electricity consumption forecasting based on clustering algorithms and fuzzy matrices in buildings", *Building Services Engineering Research and Technology*, Vol. 43, no. 16, pp. 703–724, 2022.
- [27] A. Galli, M. Savino, V. Moscato and A. Capozzoli, "Bridging the gap between complexity and interpretability of a data analytics-based process for benchmarking energy performance of buildings", *Expert System with Applications*, Vol. 15, no. 1, pp. 388–403, 2022.
- [28] M. M. Ouf, H. B. Gunay and W. O'Brien, "A method to generate design sensitive occupant-related schedules for building performance simulations", *Science and Technology for the Built Environment*, Vol. 25, no. 1, pp. 221–232, 2019.
- [29] B. Dong, D. Yan, Z. Li, Y. Jin, X. Feng et al., "Modelling occupancy and behavior for better building design and operation—a critical review", *Building Simulation*, Vol. 11, no. 4, pp. 899–921, 2018.
- [30] H. S. Park, M. Lee, H. Kang, T. Hong and J. Jeong, "Development of a new energy benchmark for improving the operational rating system of office buildings using various data-mining techniques", *Applied Energy*, Vol. 173, no. 1, pp. 225–237, 2016.
- [31] Z. Yang, J. Roth and R. K. Jain, "DUE-B: Data-driven urban energy benchmarking of buildings using recursive partitioning and stochastic frontier analysis", *Energy and Buildings*, Vol. 163, no. 1, pp. 58–69, 2018.
- [32] K. Park and S. Son, "Novel load image profile-based electricity load clustering methodology". *IEEE Access*, vol. 7, no. 1, pp. 59048–59058, 2019.
- [33] L. Wen, K. Zhou and A. Yang, "Shape-based clustering method for pattern recognition of residential electricity consumption", *Journal of Cleaner Production*, Vol. 212, no. 1, pp. 475–488, 2019.
- [34] L. G. Swan and V. I. Ugursal, "Modeling of end-use energy consumption in the residential sector: a review of modeling techniques", *Renewable Sustainability of Energy Review*, vol. 13, no. 8, pp. 1819–1835, 2009.
- [35] J. Kim, H. Naganathan, Y. Moon, O. Chong and S. Ariaratnam, "Applications of clustering and isolation forest techniques in real-time building energy-consumption data: application to LEED certified buildings", *Journal of Energy Engineering*, Vol. 143, no. 5, pp. 1–20, 2017.
- [36] H. Yassine, G. Khalida, A. Abdullah, B. Faycal and A. Abbes, "Artificial intelligence-based anomaly detection of energy consumption in buildings: a review, current trends and new perspectives", *Applied Energy*, vol. 287, no. 1, pp. 1–26, 2021.
- [37] S. Rodrigo and C. Marcelo, "Extended isolation forests for fault detection in small hydroelectric plants", *Sustainability*, vol. 12, no. 1, pp. 1–16, 2020.
- [38] A. Daniel, G. Katarina, F. Hany, A. Miriam and B. Girma, "An ensemble learning framework for anomaly detection in building energy consumption", *Energy and Buildings*, vol. 144, no. 2, pp. 191–206, 2017.

- [39] S. Jakob, T. Erik and L. Michael, "Anomaly detection forest", 24th European Conference on Artificial Intelligence, Belgium, Brussels, pp. 1–8, 2020.
- [40] H. Sahand, C. Matias and J. Robert, "Extended isolation forest with randomly oriented hyperplanes", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 1, no. 1, pp. 1–12, 2019.
- [41] B. Elhadj, D. Belkacem, I. Bachir and A. Khadidja, "Numerical simulation of conjugate convection combined with the thermal conduction using a polynomial interpolation method", *Advances in Mechanical Engineering*, Vol. 9, no. 1, pp. 1–7, 2017.
- [42] A. Abdelaziz, V. Santos and M. S. Dias, "Convolutional Neural Network With Genetic Algorithm for Predicting Energy Consumption in Public Buildings," *IEEE Access*, vol. 11, pp. 64049-64069, 2023.
- [43] H. Zhao and F. Magoules, "Feature selection for predicting building energy consumption based on statistical learning method", *Journal of Algorithms & Computational Technology*, vol. 6, no. 1, pp. 59–77, 2012.
- [44] M. Inga, "Feature selection for energy system modelling: identification of relevant time series information", *Energy and AI*, Vol. 4, no. 1, pp. 1–14, 2021.
- [45] J. Lee, J. Kim and W. Ko, "Day-ahead electric load forecasting for the residential building with a small-size dataset based on a self-organizing map and a stacking ensemble learning method", *Applied Sciences*, Vol. 9, no. 1, pp. 1–19, 2019.
- [46] A. E. Ioannou, D. Kofinas, A. Spyropoulou and C. Laspidou, "Data mining for household water consumption analysis using self-organizing maps", *European Water*, vol. 58, no. 2, pp. 443–448, 2017.
- [47] Y. Long, M. Tang and H. Liao, "Renewable energy source technology selection considering the empathetic preferences of experts in a cognitive fuzzy social participatory allocation network", *Technological Forecasting and Social Change*, vol. 58, no. 1, pp. 421–432, 2021.
- [48] A. Abdelaziz, V. Santos and S. Dias, "Machine learning techniques in the energy consumption of buildings: a systematic literature review using text mining and bibliometric analysis", *Energies*, Vol. 14, no. 1, pp. 1 - 25, 2021.
- [49] T. Räsänen, J. Ruuskanen and M. Kolehmainen, "Reducing energy consumption by using self-organizing maps to create more personalized electricity use information", *Applied Energy*, Vol. 85, no. 4, pp. 830–840, 2008.

# A Survey of Evolving Performance Analysis Technologies, Algorithms and Models for Sports

Shamala Subramaniam<sup>1</sup>, Manoj Ravi Shankar<sup>2</sup>, Azyyati Adiah Zazali<sup>3</sup>, Hong Siaw Swin<sup>4</sup>, Zarina Muhamed<sup>5</sup>,  
Sivakumar Rajagopal<sup>6</sup>, Mohamad Zamri Napiyah<sup>7</sup>, Faisal Embung<sup>8</sup>

Department of Communication Technology and Networks, Universiti Putra Malaysia, Selangor, Malaysia<sup>1,2,3,4,5</sup>  
Department of Sensor and Biomedical Technology, Vellore Institute of Technology, Vellore, India<sup>6</sup>  
Information Technology Division, National Sports Council of Malaysia, Kuala Lumpur, Malaysia<sup>7,8</sup>

**Abstract**—The emergence and extensive development and deployment of Industrial Revolution 4.0 have distinctly transformed the methodologies of sports performance monitoring. Consequently, there has been an increase in the emergence of new and adapted technologies in various areas of sports, such as competition analysis, player performance analysis and many others. There are rich and heterogeneous sports performance analysis technologies, algorithms and frameworks which provide constant basis for elevating new horizons of sports technologies. Thus, this paper aims to encompass significant findings that will provide a comprehensive survey in this area. Previous surveys have extensively focused on various methodologies of sports performance analysis, sport-specific analysis and other technology revolving around sports performance analysis. However, most of the focus is largely on training and competition performances and not off-field. The objective of this paper is to understand the current research trends, challenges and future directions of dynamically evolving technology embedded in the world of sports. This survey aims at contributing to this rich repository but with a new focus element of off-field that researches the connection between the athlete, the sports aspect of their life, the non-sport aspect and the methodologies of sports performance analysis. In addition, the exponential growth of Artificial Intelligence (AI) as a base for sports performance analysis systems and platforms is analysed extensively. This paper also presents a comprehensive classification of athlete performance analysis using algorithm tools and sports performance platforms and systems. Subsequently, the detailed analysis of this taxonomy has enabled the identification and detailed analysis of open issues and future directions.

**Keywords**—*Sports performance analysis technology; on-field analysis; IoT; real-time monitoring; off-field analysis*

## I. INTRODUCTION

The Industry Revolution 4.0 (IR 4.0) has seen the extensive harnessing of multitudes of technologies in wide and rich spectrum of areas [1]. This encompasses the domain of sports and is a significant element in positioning itself as a national and global agenda. Since the inception of technological advances, the sporting world and its entities have been strengthened in multiple aspects [2], such as the usage of wearable Global Positioning System (GPS), sensor technology, virtual imaging, Hawk-Eye Line-Calling System, and time tracking systems. These contributions to sports have increased the accuracy of measuring equipment and instruments. Among the reference success cases of sportsmen partnering with

technologist are like athletes Kell Brook have worked with Sheffield Hallam University in the lead-up to his International Boxing Federation (IBF) world welterweight title fight. The scientists collected Heart Rate (HR) and lactate data. Like Brook, multiple-time National Basketball Association Most Valuable Player (NBA MVP) Stephen Curry has overcome his physical shortcomings by incorporating technology into his training regime. The All-Star basketball player has used strobe goggles and on-court light discs that force a sensory overload and demand quick decisions. Team sports athletes constantly strive to improve their international and league rankings. Football athletes in the second division of England, Germany, and France have been monitored with over 11,000 team-matches observations to monitor the factors influencing the chance of promotion to the elite leagues [3]. The researchers executed a series of logistical regression analyses and made observations, proving that teams will do anything and everything to evolve and improve their status in the sporting world constantly. The athletes and coaching staff of the modern generation have resourced technology tremendously to enhance their athletic ability [1]. The cases presented above serves as a basis to justify the need to articulate the rich repository of sports technology to enable the trend analysis and determination of open issues especially with the rise of Artificial Intelligence (AI). The analysis of off-field and on-field correlations from a distinct perspective of existing work has high constraints. Thus, creating problem in identifying pertinent open issues in this domain of analysis. Thus, this survey paper is objectives are to address the solutions for this problem. The goal of this survey paper is to show the state-of-the-art sports performance analysis technologies. It encompasses a comprehensive review of papers. In the papers we reviewed, each algorithm, architecture or system is reviewed in detail from the implementation, their advantages and respective disadvantages. The developed taxonomy of comparison has provided a detail basis of the identification of the open issues.

This paper is organized as follows. Section I is the Introduction. Section II discusses in detail the various Surveys that have been published in this area and highlights the uniqueness of our Survey. Section III presents the Proposed Classification Model based on the research from three subsections: Competitive and Training Performance Monitoring or On-Field, Non-competitive Performance Monitoring or Off-Field, and Systems and Platform. Open

issues collectively analysed throughout this paper, recorded and arranged in relevance discussed in Section IV with six subsections: Relating the Monitored Research of Athletes on Competitive and Training and the Non-Competitive States, Adapting to Sports Performance Monitoring Systems, Combating Athletes' Stress, Rise of Extensive research in Machine Learning (ML) and AI Will Enable the Dominance of Demographics, Sport-Specific, and Data-Handling Ethics. Finally, the paper concluded in Section V.

## II. ANALYSIS OF RELATED SURVEYS AND TAXANOMIES

Reiterating the important fact that the recent acceleration in sports technology has motivated researchers towards the inclination to publish research based on sports performance analysis. This section discusses and reviews extensively the previous surveys conducted on sports performance analysis and the critical issues surrounding them. Research in [4] surveys elite and pre-elite athletes, evaluating unseen factors contributing to their success. 135 Australian Olympic, Paralympic, National, and state-level athletes from 25 Olympic sports were surveyed. Our research has also improved and elaborated that there are more factors than those included in athlete development programs such as lifestyle, social and support factors. In [4] it has been found that international athletes perceived psychological skills and attributes, along with strong interpersonal relationships, as vital to their success, and they also rated 'Recovery practices' as very important and made extensive use of available support services. However, the athletes have indicated the necessity for access to these services at the grassroots level. The study has concluded that athlete development systems need a complete environment that allows athletes to succeed, perform consistently, have longer careers, and gracefully transition into retired athletes.

In [5] the research done on analysis methods in sport for intelligent data has been reviewed in detail. More than 100 studies on intelligent data and its analysis methods use Smart Sport Training (SST). Some of the methods among others surveyed included Computational Intelligence (CI) methods like fuzzy systems and simulated annealing, data mining methods like Support Vector Machines (SVM), and Random Forests (RF), Deep Learning (DL) methods like Recurrent Neural Networks (RNN) and Conventional Neural Networks (CNN), and other methods like Naive Bayes (NB) and Bayesian Networks (BN). The research also classified the research surveyed by sport type; individual, mixed, and team. Researchers have focused their attention on soccer, running, and weight lifting. The relation to participation levels, over half of the research study focused on individual sports, with team and mixed sports accounting for a third of the total. The research elaborated on the study type done on a particular sport and the focus of the research and results. This research may improve by adding more validation-level research publicly available with the datasets for replicating research, which improves methods.

In team sports, numerous variables influence the outcome and performance of the teams. The research in [6] surveys team sports and the usage, challenges, and techniques of implementing AI and ML with computation, including forecasting match results, tactical decision making, player

investments, fantasy sports, and injury prediction. The work evaluation on match outcome prediction found that, due to the unpredictability of sports, models still fail to forecast outcomes much better than bookmakers and appear to have hit a barrier, but there are several feasible solutions. This article also demonstrated the possibility of developing a one-of-a-kind real-world live testbed for AI and ML approaches to be validated in the future. According to a literature survey in the fantasy sports area, there are some AI approaches in the Fantasy Premier League (FPL) football competition to beat most human players dramatically. Overall, this study illustrates the impact of AI and ML approaches on the team sports domain, highlighting some processes with open areas and research issues. The survey focuses the research on six sports where only a finite amount of literature has been done. The narrow scope has allowed the team to be fortunate in finding research that contains the highest accuracy, cricket, with 75% in the prediction model. A broader scope would have seen sports with far more uncertainties and lower levels of accuracy in their research.

In [7] the role of ML in predicting and avoiding sports injuries has been discovered. The article uses Tree-based ensemble methods, SVM, and Artificial Neural Networks (ANN) ML methods. Pre-processing steps aided the classification algorithms, enhanced over and under sampling methods, hyper parameter tuning, feature selection, and dimensionality reduction. The comprehensive study found that ML technologies forecast sports injuries in 11 researches. The study closes by requiring ML to identify high injury risk athletes and essential injury risk indicators. AI offers a fascinating new viewpoint on injury risk and team sports performance prediction. Another literature study [8] covers AI in sports medicine, data processing, injury diagnosis, and prevention in competitive sports. Models and approaches in the literature study include fuzzy sets, ANN, Markov process, and other models such as Bayesian theory and multi-dimensional models groups. The review needs to be more systematic, a weakness of the research. In [9] the implementation of GPS units to collect data on a full-time basis has been done. The athletes, 52 players, enrolled into the Korean National Team, provided data to calculate the optimal ratio of Acute to Chronic Workloads (ACWR). The observational study reviewed other injury related research to deduce the calculation behind workload and the probability of an injury occurring to an athlete. Unlike the previous studies, which focused on pre-collected data, the research quantifies the workload of 52 athletes using GPS units collected during game-based training and matches. The research has filled the need for a standalone study, which has also conclusively suggested that hockey athletes and their ACWR should stay within the moderate low, especially for strikers and midfield playing positions, to manage non-contact and soft tissue injury.

In [10] the functional usage of ML and CI in sports prediction has been surveyed. ML, ANNs, BN and Logistic Regression Methods, SVM, and Fuzzy Logic and fuzzy systems are some models discussed in sports-related works. Many elements influence the outcome of a sporting event, including a teams' (or a players') morale, talents, and coaching plan. This review paper examines past research on data mining



methods for predicting sports outcomes and weighs the benefits and drawbacks of each approach. However, some sports, such as most track and field sports, are simply too easy to justify the complicated framework to the point where it is no longer necessary. There is a use of deep NN approaches in team sports analytics. Sports analytics using a DL approach is now possible thanks to tracking and visual data in sports and recent technological advancements. In [11] the use of modern DL techniques in team sports analytics has been reviewed. The survey researches two sports among the team sports that have benefitted from sports analytics, basketball, and football. The survey has tracked the advances in DL techniques in the two sports. The researchers aim to provide a study that provides insight to team sports analysts in sports and the ML aspect.

As discussed in the performance analysis subsection, [12] surveys the journal databases, reviewing the literature on Smart Wearables. The research classifies health, sports, daily activity, tracking and localization, and safety into four major clusters. However, data resolution of wearable sensors, power consumption, wearability, safety, security, regulation, and privacy became the primary obstacles of wearable Internet of Things (IoT) devices. In [13] the need for performance analysts within the coaching process within elite football coaches has been evaluated. The research dissects the differences in the necessity of PA's at the professional and football academy levels. The purpose of this study was to fill a gap in the literature on the function of match analysts in providing feedback via match and notational analysis techniques and systems. The exploratory study uses an online questionnaire based on information from current match analysts in elite football, academic practitioners in performance analysis, and current literature. 48 match analyst practitioners from significant football clubs completed the survey. The majority of 32 analysts worked in a professional team setting, while 16 worked in an educational setting. Educators and coaches can use the data gathered from training sessions and games analysis to understand better the challenges faced by a trainee, a player, or even an entire team and develop appropriate training and strategy plans. The research done by [14] has heavily influenced the structure and taxonomy due to emphasis on deliberating the finding of current surveys, current techniques, and trends of performance monitoring. This paper then proposes a classification scheme for these systems, separating them into invasive and non-intrusive categories. Researchers prefer nonintrusive systems since they do not interfere with the game. Each system's unique traits and strengths and weaknesses are listed. However, the system is still early and cannot extract high-level metrics such as game circumstances, team formations, or psychological characteristics.

The discussion in [15] is an in-depth understanding of content-aware systems for sports video analysis by examining the insight offered by research into the content structure under different scenarios. Themes relevant to the research on context-aware systems for broadcast sports were analysed. Analysis can benefit significantly from the use of ML. After evaluating coaches' responses, the study found that the system is valuable to daily work. The research summarizes the future trends and challenges for sports video analysis and sets the tone for the rest of this study in the section on video analysis. On the other

hand, developing a unified framework that enables processing data from diverse sports is still challenging. The trade-off between commonality and robustness must prevail because the future goal of action recognition in sports is to develop a machine that can read, write, listen to, and speak a voice over to broadcast sports videos directly.

The uniqueness of creation is that no two human beings are the same. Similarly, the conditions of individuals are different based on their fitness level and training consistency [16]. The researcher must determine a standard or baseline for every athlete individually to evaluate the athlete's performance. This way, it is considered that all humans are unique and may react differently to the stimulus applied, thus increasing the accuracy of studies and making surveys more accurate for the reader's comprehension in mapping them to specific domains. In contrast, more requiring research on Tai Chi and Qigong effects has only a few studies. Future research could also use shorter time intervals between RHR measurements to understand the underlying processes, potentially contributing to RHR decreases. The research done by [17] gave an in-depth method of monitoring an athlete's sports performance or a team of athletes, and this includes looking at several potential moderator variables' effects, and the cohesion-performance relationship revealed in research utilizing the Group Environment Questionnaire (GEQ). Standard literature searches turned in 46 studies with 164 effect sizes in total. This analysis breaks down the cohesiveness and performance in the sport. The GEQ had a moderate effect in studies that employed it. Refereed publications (as opposed to unpublished sources) and female teams had a more substantial cohesion performance effect. The research further breaks down the methods of measuring splitting variables like gender, type of sport, level of skill or experience of the athlete, and data source.

Face video-based Photoplethysmographic (PPG) signals acquired with professional or consumer-level cameras to obtain HR remotely. In [18] the latest advances in video-based HR management were surveyed. The research focused on the technological updates that overcame the existing and overwhelming challenges caused by illumination variations and motion artifacts. The majority of available remote Photoplethysmographic (rPPG) methods currently work with uncompressed video data. Conversely, the uncompressed videos will take up a lot of disc space, making internet data exchange impossible. The background of imaging Photoplethysmographic (iPPG) and rPPG, which is an estimation method for HR, was discussed, and debating the prospects of this technique and potential research direction in [19]. PPG and noise reduction using wavelet transforms to measure people's HR, which proposed recreation of the method for obtaining HR from the rPPG. rPPG is a technology that uses current or previously recorded video from a simple web camera to estimate HR, oxygen saturation, and other parameters. The heartbeat is usually highly regular over a short time; these physiological characteristics estimate that arteries blood flow shows some periodic flow. As a result, slight fluctuations in the amount of light reflected from the face are visible in the arteries and blood vessels of the face, which can be caught by the camera and processed as a Blind Source Separation (BSS) problem. The research successfully created a

real-time system that detects an individual's face and facial tracking and displays the HR with maximum noise reduction. The work could be improved by incorporating the ML technique into the PPG signal identification and increasing face detection accuracy.

HRV is a promising and essential research technique for cardiovascular disease diagnosis. The Parasympathetic Nervous System (PNS) and Sympathetic Nervous System (SNS) of the Autonomic Modulate System (AMS) regulate and control the HRV. HRV analysis can evaluate a variety of cardiological and non-cardiological illnesses. The research done by [20] surveyed HRV and the linear methods involved in the methodological evaluation. The two linear domains are the time domain and frequency domain. The researchers also discussed nonlinear methods of HRV like Poincare Plot Analysis, Approximate Entropy (APEN), Sample Entropy (Sampentropy), Detrended Fluctuation Analysis, and Correlation Dimensions. The parasympathetic and sympathetic controls, on the other hand, may have an impact on the alpha value in the study and fail to discriminate between them ultimately. As a result, a separate examination of both the short-term and long-term scales is necessary to determine the actual range of the scale as it withdraws the reciprocal effect.

In [21] the focus is on football as the ML applications in sports analytics relate to player injury prediction and prevention, potential skill, or market value evaluation. CI has shown to be a valuable tool in various fields. This study looks into the possibility of predicting long-term team and player performance. By surveying 31 categories of study and deriving the methodologies, information, and applications, large amounts of data turn into meaningful knowledge through data mining. Historical data and advanced statistics offer a reliable projection of the final league table and whether a team will have a more robust season. The findings taken from different leagues show a significant disparity. As a result, essential differences across leagues should apply universally. Player exhaustion and severe long-term injuries are problems that can harm players or teams, but if addressing the intricacy of the situation, such data could be valuable study tools. Many sports organizations have begun to understand that the data previously retrieved has a treasure of undiscovered knowledge, as data mining techniques capture the attention of the information industry and society due to a significant volume of data and the impending need to turn it into valuable knowledge. The research in [22] classified the 31 articles into nine thematic types with a systematic review of sports data mining from 2010 to 2018. The researchers also located possible areas to be explored, such as swimming, athletics, hockey, boxing, fencing, and tennis. The review concluded the survey by encouraging new research in this field.

The work in [23] surveyed multiple online databases for articles using AI techniques applied to the team sports athletes. The team applies AI to predict injury risk and team sports performance possibly. The most used methods surveyed are ANN, decision tree classifier, SVM, and Markov process, including good performance metrics. Soccer or football, basketball, handball, and volleyball were researched as the traditional team sports. The research concluded with the assurance of a promising AI and team sports future. There are

some differences in sample sizes in the manuscripts, with some samples being lesser than others.

The review done by [24] analyzed the literature on sports predictions that have utilized the application of ML. The research categorizes the method of ML into unsupervised learning and reinforced learning. Location, player health, player performance, weather, and ground conditions are all factors that influence whether a game is won or lost. Plenty of data is available for long-seasoned and high-scoring games like basketball, making prediction considerably more manageable, but guessing the outcome for games that are only played once a year and are low-scoring becomes a problematic endeavor. The team then attempts to show the comparison using a table, displaying the approach, game, technique, and review of the research surveyed. The research concludes by highlighting the study's limitations and prospects.

In [25] the microsensors usage and the monitoring approach implemented in basketball and did an online survey, and applied multiple responses, Likert-scale level of agreement, and open-ended questions on basketball practitioners were researched. Questions for the basketball practitioners included how player monitoring was performed, highlighting the barriers and facilitators with microsensors. Nearly two-thirds of respondents implement player monitoring, and almost one-third of basketball practitioners use microsensors. The survey concludes that basketball has low uptake of microsensors in sports performance monitoring. Because of this study's small sample size, it was impossible to analyse results based on criteria such as the playing skill of respondents.

The study in [26] did a survey on recovery strategies among basketball practitioners. The majority agreed that recovery strategies are very vital in their routine. The best strategies were active recovery, massage, foam rolling and stretching. The biggest challenges for the basketball athletes surveyed would be the lack of devices and facilities, high cost and lack of time. The research does find that there is a disassociation between scientific evidence and perceived evidence. The survey also noticed the inclination of athletes to prefer easily implemented strategies rather than evidence-supported strategies.

The work in [27] has given a comprehensive explanation about wearable monitoring systems. There are highlights of the usage of sensors, specifically commercial wearable systems for sports applications. The book overviews the psychological parameters measured by wearable systems. The HR and oxygen consumption parameters are analyzed in the respective sections. The following sections discuss the practitioners and the sports utilizing these wearable systems. The book concludes by evaluating the immense value of wearable systems to multiple sectors, including sports. The authors express their opinion on the current business expansion of wearables, which is currently small, and market forecasting has yet to produce an accurate and generally disseminated insight. Wearables technology can collect rich contextual data from the device itself and use it to provide a truly tailored experience.

In [28] the work surveys wearable technology, the progression in the development of wearable devices, and the

latest advances in the wearable devices market. They also classified the wearable devices based on factors and analysed in-depth information on the technology, highlighting the adverse challenges and prospects of wearable devices. Due to the lack of good practices in interoperability and proper standardization in the new Internet of Wearable Things (IoWT) niche, the close connection of various systems provided by different suppliers remains one of the most critical issues of wearables. AI in Sports Performance, ML and Sports monitoring, and RTM are the primary demographics and focus searched for throughout the research.

### III. PROPOSED CLASSIFICATION MODELS

In this paper, we proposed two classification models as a basis to address the importance of the term “off-field” performance analysis tool. This proposed classification being given a distinction is based on the extensive survey conducted in the previous section and the analysis has inspired that there is a need to give the analysis of athletes beyond the training and competition venue. Thus, creating the need to empirically analyse their off-field activities and their impact on the on-field. In this section we will review pertinent research done in off-field and subsequently the next section will present the open issues we have identified based on our extensive survey.

The ecosystem of an athlete consists of a broad spectrum of activities with multiple interactions with varying groups of people. This research surveyed 50 athletes, from national to state, and added the most common results.

Different athletes may have different timetables or schedules. Top-tier athletes train at least one or more times a day. Before stepping on the pitch, many athletes have Pregame Rituals (PGR) to complete [29]. In Ghana and other Sub-Saharan African countries, unorthodox PGR is common in sports, mainly soccer. Based on scientific descriptions, empirical investigations, and specific field observations, the paper also underlines the conceptual contrasts between PGR and pre-performance routines. The study in [30] put these superstitions to the test among track and field athletes, where there is more to the athlete’s preparation than mere superstition. The positive correlation between athletic identity and superstitious behaviour shows that student-athletes with strong athletic identities used more superstition in sports events. Individuals with a high athletic identity utilize superstition as a coping mechanism to minimize anxiety during the competition [31] and protect their egos. Precompetitive Mood States (POMS) have been reviewed by [32] and give us the ideology behind the Mental Health Model of the athlete before a competition as well as the more suitable Hanin Individual Zone of Optimal Function (IZOF) model. An athlete must consider all facets of the preparation time allocated before the tournaments. Whether on the pitch or off the pitch, every move imprints the athletes. There are multiple studies targeted at the effects of the actions carried out by athletes when the athletes are in a non-competitive state. The research surveys the studies devoted to uncovering the truth behind the effects of non-competitive actions on the athlete’s performance.

The proposed patent Prest and Hoellwarth is looking to monitor the vital signs of an athlete using headphones, earbuds, or headsets. While managing the electronic device, the

monitoring system monitors user activity during exercise or sporting activities. Other user characteristics such as biometric data, temperature, sweat, and HR are attributable to the monitoring system’s placement. The usage of headphones or earphones is rising during training sessions, despite having multiple side effects such as perforated eardrums [33]. There has been a study relating the effects of music on exercise [34]. It is also viable for a non-competitive approach. The athlete might want to calm his nerves before and after the game, and during that period, a reading of the HR, perspiration level, and steps taken should be recorded. Music can have a considerable positive impact on exercisers and athletes, especially in terms of increased effective reactions and physical performance, reduced perceived exertion, and more efficient oxygen consumption. The duration an athlete takes to reach a state of physical readiness for the next activity allows coaches to plan out substitutions of athletes for the games. Other alternatives like monitoring the step counter through the shoes have lost some avenues to monitor vital signs like HR and perspiration levels as effectively as this patent. The research in [35] measures an athlete’s performance without measuring their physical performance, but instead taking a psychological measurement, Sports Performance Inventory (SPI). A principal components’ analysis with Varimax Rotation performed on the original survey items resulted in an 83-item survey with six interpretable factors: competitiveness, team orientation, mental toughness, emotional control, positive attitude, and safety consciousness. Compared to novice athletes, college athletes had a higher SPI composite, a more positive attitude, and were more competitive. Females were more team-oriented than men, and novice males were more competitive than novice females, with college females outperforming college males. However, there is no direct proof that the SPI’s dimensions will predict an athlete’s performance, and sex differences discovered between athletes may be premature due to the small sample size.

The health industry has a huge responsibility. Discharged patients are still under the monitoring eye of the hospitals. The common issues for missing monthly check-ups are distance and transportation. The analysis in [36] has multiple uses and currently monitors the patients’ vital signs. The developed VJ is a microelectronics and textiles based vital signs monitoring system. The VJ has multiple readings for the usual ECG, respiration, perspiration, and oxygen saturation percentage. One unique feature that a clothing-based sensor can detect is posture. The VJ can detect the posture reading of the test subject. The posture reading can be beneficial in a sports-based environment because sports require posture perfection, which could be one of the solutions for these sports. Moreover, games that find an external wearable sensor disrupt the gameplay’s efficiency. Not to mention the question of safety, some sensors are waist-mounted, and even watch-based can quickly disrupt a play of games like rugby, whereas a shirt or vest used under the players’ uniform will form minimal resistance for the gameplay.

Other studies have touched on the negative behaviours among athletes and how that affects training to gather evidence that no matter to what extent the athlete puts the body through gruelling training sessions, the effects from non-competitive

activities leave a lasting impression on the performance. Drugs are categorized by [37] into usage and abuse [38]. The usage of allowed substances for an athlete's recovery has long been a part of the life and diet of performance athletes, but there is a fine line between usage and abuse. The overall harm of drug misuse comprised the substance's direct physical harm to the individual user, the drug's ability to create dependence, and the impact of drug abuse on families, communities, and society [39] [40]. Marijuana addicts are more likely to develop a persistent cough, bronchitis, and lung and upper airway cancers. Regular marijuana usage has sadness, anxiety, and schizophreniform illness in some people with a pre-existing predisposition [41]. Many drugs are outside the Health Ministry and International Sports governing bodies such as the World Anti-Doping Agency (WADA) [42]. The study in [43] highlights drugs like cocaine, meth/amphetamine, ketamine, and other drugs among "full-time" athletes and discusses the effects on the population of "full-time" athletes. The monitoring happens on a self-reporting basis and drug tests. The survey has also clarified that drugs harm current performing athletes and affect them when they stop competing and retire. It is common knowledge to note that drugs are harmful to any class of humans, but as the study has thematically targeted athletes, the survey discovers that athletes are no exception.

The athlete must recognize the body's biological signs and capabilities and sense what the body is going through. The research in [44] monitors the point at which the subject has reached the level of fatigue. This analysis documents a self-reporting, and the athlete notes the point at which he or she reaches fatigue while playing a sport of the choice. The test subjects needed to complete three surveys daily, and some relied solely on these surveys to convey what the test subjects were experiencing. The test subject saw that as the weeks progressed, there was more and more load that test subjects could handle, and this showed that progressively overloading the body meant that the body was less likely to feel fatigued at an earlier stage. The body muscles get accustomed to the athletes' load, and as the documentation proved, the athletes can withstand more the next time the test subjects are under the tests.

In [45] the manner in which an athlete's training affects their competitive performance is examined. The researchers analyze the link between competitive disc throwing performance and maximum lifting weights in female disc throwers. Maximum lift weights were recorded for the bench press, full squat, deadlift, high clean, and snatch. They use Pearson's R Accumulated Correlation Coefficient to determine the relationship between competition performance and 1-RM. Weights show a substantial positive link between female disc throwers' performance and their maximal lifting weight in the bench press, high clean, and snatch as female disc throwers weigh less than male disc throwers; hence they need to throw faster to convey the same amount of force. The high clean and snatch actions may contribute to power output during the delivery phase.

The COVID-19 pandemic is an example of an extended period forcing the athlete to be away from training. Almost all nations enforced nationwide lockdowns. The research in [46]

observes the usage of Virtual Reality (VR), which has already proven to be a step forward compared to video playback training. VR was used to separate the visual data of player movements from the visual data of the ball trajectory. The three conditions described are the player's throwing action, ball trajectory, and final location. The immersive environment emulates the field of play, the players, and the game methodologies, allowing for a comprehensive game mode to be tapped and trained in the athlete.

#### IV. OPEN ISSUES

The extensive discussion and review of the wide spectrum of research conducted in off-field and on-field has distinctly and empirically illustrated the depth and spectrum of research in sports performance analysis. Our analysis has derived the following open issues which will further enhance the spectrum of harnessing resources *to elevate sports*.

##### A. *Relating the Monitored Research of Athletes on Competitive and Training and Non-Competitive States*

In summary, when writing this survey, there is no predefined research between the stress applied to the athletes from the non-competitive daily routines and the performance displayed during competitive and training states. Studies performed on monitoring an athlete in competitive and training states, creating baseline readings, and through training and practice comparing the athlete's current state and performance over a while. Despite having studies done to relate pre-competition state using POMS, and the athlete's performance during competition, no specific correlation has been derived between what happens when the athlete leaves training or competition. An athlete spends an average of 20 to 30 hours a week in training and doing on-field activities, and there is an average of 140 hours that an athlete is away from the field. Therefore, 80% of the athlete's schedule is bound to affect the 20% spent on the field.

##### B. *Adapting to Sport Performance Monitoring Systems*

As sporting fraternities continue to evolve, all avenues must ensure excellence. An avenue like a stress monitoring system to access the athlete's current state will be extremely valuable and open to being tapped. There has been a void in this sector that has limitless potential in the case of pursuit. The sports world must accept the digitalization of performance analysis. Newer and more advanced systems than video monitoring have emerged in the sports fraternity. A recent survey researched the perception of rink-hockey head coaches and the usage of performance analysis as a tool to assist training, match preparation, observation, and interventions. The research has further cemented the importance of performance analysis by including seven experienced First Division Portuguese rink-hockey head coaches and conducting semi-structured questions, and the data analyses through inductive and deductive content analyses. Rink-hockey head coaches prefer to analyse the opponents themselves to plan training, assist with tactical preparation, and implement within-match strategies. They considered video analysis a vital tool to analyse opponents' strengths and weaknesses, focusing on the opponent's goalkeeper. Rink hockey has adopted performance analysis to prepare for tournaments and world sporting events in their armoury, like many other sports. Like rink hockey,

many other sports and coaching staff must be open to working with researchers, providing data on their athletes, and taking part in performance analysis of the individual athlete or team. The grassroots federations have less to worry about introducing new technology because the severity of the contracts and agreements, if present, is less than that of the professional leagues. The athletes are younger and will be more open to accepting and working with new technology introduced by the researchers. Also, wearable devices in grassroots level training sessions and games will not hinder high-profile games, unlike the professional games played by star athletes.

### C. Combating Athletes' Stress

An athlete is the face of the country in international multi-sport events, and there should be no unnecessary stimuli that bother the athlete. Research on the lifestyle of an athlete analysis the action that may induce stress recorded through HRV and isolates the athletes from such actions. An athlete should be able to allow competing without stress, which in turn is known to cause performance anxiety. Research that proves the importance of athletes' surroundings and the effects on performance will encourage national sports bodies to shape the surroundings of athletes according to their needs and remove stress triggers to ensure more potential tapped out of the athlete and more glory brought to the nation.

### D. Rise of Extensive Research in ML and AI will Enable the Dominance of Demographics

The modern sport of hockey boasts a fast-paced, physically grueling sporting event. One such sport is field hockey; with over two billion viewers annually and the top five most watched sports globally, minimal documented studies have been done on the sport or its athletes. Athletes of the top calibre require very high levels of stamina. Elite athletes need to maintain a high fitness level and constantly train their skills. The skills of hockey are displayed in the accuracy of shots and passes, the ability to run with the ball, and take power shots. The athletes in the sport need to monitor their performance during the competition and training and non-competitive states if the aim is to reach the top level of hockey play. The National Level hockey players could benefit from the study and the knowledge that many activities knowingly or unknowingly cause the dip in sports performance during athletes' competition and training. It will be another tool in the arsenal for the coaches and athletes to exploit and further improve.

### E. Sport-specific

The study has surveyed an array of different sports and the method of monitoring athletes and monitoring the performance of the athletes. The difference from one sport to the other changes drastically the more profound the research goes into the details and skills of a sport. Every sport has a unique skill set and abilities that the athletes must complete, and sports monitoring has been very sports-specific. Some sports have a huge reception globally, but very little documented research. Sport-specific research proposed directing towards that is focused on the uniqueness of the sports and the respective demographics. The research should cover an existential issue of the effects of the athletes' non-competitive practices and how top-level athletes are affected by it. The study should cover the area of wearable sensors as their go-to research

method. Using the current systems in the market, like Suunto, boast a repertoire of precision and accuracy in measuring the physiological changes in the body. Unfortunately, mega-corporations like Suunto do not divulge the readings and information recorded on their watches to the public or is search team. They expect users to use their hardware and readings taken at face value. Another option is the Samsung Watch due to their operating system running on Tizen. Tizen allows developers and researchers to develop their applications to be tested in the Samsung Watch hardware, applying using the sensors optimally. The Tizen software also allows for the readings to be taken, recorded, and analysed by researchers.

### F. Data-Handling Ethics

Bio-metric data has been a repeating element in most of the research discussed in the survey. The power and accessibility of the technology behind bio-metrics have made it very susceptible to monetization or the user's identity being stolen. In the professional sporting sector, some laws involve data sharing policies. The agreements and contracts for athletes allow their documents and data to be protected from being misused, which is not the case for lower levels of the sport. There is a gray area in the confidentiality of athlete data collected by researchers. There has to be a standard protocol for all athletes and researchers, allowing for more ethical practice in sports research. The current trend leaves a gap that hackers can exploit.

## V. CONCLUSIONS AND FUTURE WORK

This paper presents a survey on the sports performance analysis of athletes' competitive and training states in various sports. We also surveyed the athletes and monitored their non-competitive activities or away from training. Then taxonomy of comparison was done on the research based on metrics like using the wearable sensors. The most common method in the modern age is to use mobile wearable sensors on the athletes while monitored during competition and training. However, there is also an avenue to use the same concept of wearable sensors to monitor athletes that leave the pitch. As mentioned above, an athlete's job is continuous and not only during competition and training. Therefore, we have presented the option of a bridge where non-competitive monitoring is the void many researchers may need to exploit in the open issues. On the topic of open issues, we have also found that:

- Stress monitoring among athletes is done extensively during competitive and training activities, but not once the athlete leaves training. The stress subjected to the athletes could be why the player's progress and performance are stunted.
- The national sporting bodies in nations need to pay heed to the lifestyle of athletes and foster an environment conducive to the athlete, besides the vigorous and state-of-the-art training facilities.

Sports with a fan following and stadium ambience creation by these fans need more research on their players and the sporting environment. The future direction of this research will be to encourage more researchers to create holistic and complete performance monitoring systems. The performance monitoring system or architecture must contain a registration

module for the national-level database of athletes, which acts as a recommend system for national-level athlete selection. The athlete selection must be based on collective performance data and not only a single qualifying event win. Single event qualifying win has been the method of qualifying for state to national level athletes. The lack of a complete database of athletes requires sports federations to select athletes through events, not merit or potential performance. Thus, performance analysis tools will constantly need to be enhanced and to be designed to furnish the specific and demographics in precise.

The future work will be focused on the analysis of sports performance analysis based on sports specifications. The harnessing of intelligence in its wide spectrum will be further covered. The detailed review and analysis of hardware development such as biomedical sensors is also being pursued.

#### ACKNOWLEDGMENT

This project has been supported by the Contract Research Grant of University Putra Malaysia – National Sports Council for the iGames (Vot Number - 6300375) and the UPM Innohub Market Validation Grant for IoSRocks – Mobile Device Application (Copyright (CRLY2022W04194 )) (Vot Number - 9005003).

#### REFERENCES

- [1] Mali, N.P., Dey, S.K., 2020. Modern technology and sports performance: An overview. *International Journal of Physiology, Nutrition and Physical Education* 5, 212–216.
- [2] Vanessa, R., 2020. Sport technology: A commentary. *The Journal of High Technology Management Research* 31, 1–6.
- [3] Jamil, M., Liu, H., Phatak, A., Memmert, D., 2021. An investigation identifying which key performance indicators influence the chances of promotion to the elite leagues in professional European football. *International Journal of Performance Analysis in Sport* 21, 641–650.
- [4] Burns, L., Weissensteiner, J., Cohen, M., Bird, S., 2022. A survey of elite and pre-elite athletes' perceptions of key support, lifestyle and performance factors. *BMC Sports Science, Medicine and Rehabilitation* 14, 1–12.
- [5] Rajšp, A., I., F.J., 2020. A Systematic Literature Review of Intelligent Data Analysis Methods for Smart Sport Training. *Applied Sciences* 10.
- [6] Beal, R., Norman, T.J., Ramchurn, S.D., 2019. Artificial intelligence for team sports: a survey. *The Knowledge Engineering Review* 34, e28.
- [7] Eetvelde, H.V., Mendonça, L.D., Ley, C., Seil, R., Tischler, T., 2021. Machine learning methods in sport injury prediction and prevention: a systematic review. *Journal of Experimental Orthopaedics* 8, 1–15.
- [8] Poulos, P., Serlis, A., Groumpos, P.P., Gliatis, I., 2021. Artificial intelligence and data processing in injury diagnosis and prevention in competitive sports: A literature review. *"MOJ Orthopedics & Rheumatology"* 13, 34–37
- [9] Kim, T., Park, J.C., Park, J.M., Choi, H., 2021. Optimal relative workload for managing low-injury risk in lower extremities of female field hockey players: A retrospective observational study. *Medicine* 100, 1–6.
- [10] Langaroudi, M.K., Yamaghani, M., 2019. Sports Result Prediction Based on Machine Learning and Computational Intelligence Approaches: A Survey. *Journal of Advances in Computer Engineering and Technology* 5, 27–36.
- [11] Mgaya, G.B., Liu, H., Zhang, B., 2020. A Survey on Applications of Modern Deep Learning Techniques in Team Sports Analytics, in: Abraham, A., Ohsawa, Y., Gandhi, N., Jabbar, M.A., Haqiq, A., McLoone, S., Issac, B. (Eds.), 12th International Conference on Soft Computing and Pattern Recognition, SoCPaR 2020, Springer International Publishing, Cham. pp. 34–443.
- [12] Dian, F.J., Vahidnia, R., Rahmati, A., 2020. Wearables and the Internet of Things (IoT), Applications, Opportunities, and Challenges: A Survey. *IEEE Access* 8, 69200–69211
- [13] Wright, C., Atkins, S., Jones, B., Todd, J., 2013. The role of performance analysts within the coaching process: Performance Analysts Survey 'The role of performance analysts in elite football club settings'. *International Journal of Performance Analysis in Sport* 13, 240–261.
- [14] Santiago, C.B., Sousa, A., Estriga, M.L., Reis, L.P., Lames, M., 2010. Survey on team tracking techniques applied to sports, in: *International Conference on Autonomous and Intelligent Systems, AIS 2010*, pp. 1–6.
- [15] Shih, H.C., 2018. A Survey of Content-Aware Video Analysis for Sports. *IEEE Transactions on Circuits and Systems for Video Technology* 28, 1212–1231.
- [16] Reimers, A.K., Knapp, G., Reim, C.D., 2018. Effects of Exercise on the Resting Heart Rate: A Systematic Review and Meta-Analysis of Interventional Studies. *Journal of Clinical Medicine* 7, 503–533.
- [17] Carron, A.V., Colman, M.M., Wheeler, J., Stevens, D., 2002. Cohesion and Performance in Sport: A Meta Analysis. *Journal of Sport and Exercise Psychology* 24, 168–188.
- [18] Chen, X., Cheng, J., Song, R., Liu, Y., Ward, R., Wang, Z.J., 2019. VideoBased Heart Rate Measurement: Recent Advances and Future Prospects. *IEEE Transactions on Instrumentation and Measurement* 68, 3600–3615.
- [19] Sujathakumari, B.A., Shreeharsha, B.S., Verma, P., Shivram, S., Raksha, A.R., 2018. Heart Rate Measurement using Face Video with Noise Suppression, in: 4th International Conference for Convergence in Technology, I2CT 2018, pp. 1–7.
- [20] Londhe, A.N., Atulkar, M., 2019. Heart Rate Variability: A Methodological Survey, in: *International Conference on Intelligent Sustainable Systems, ICISS 2019*, pp. 57–63.
- [21] Pantzalis, V.C., Tjortjis, C., 2020. Sports Analytics for Football League Table and Player Performance Prediction, in: 11th International Conference on Information, Intelligence, Systems and Applications, IISA 2020, pp. 1–8.
- [22] Bonidia, R.P., Rodrigues, L.A.L., Avila-Santos, A.P., Sanches, D.S., Brancher, J.D., Mustapha, A., 2018. Computational Intelligence in Sports: A Systematic Literature Review. *Advances in Human-Computer Interaction* 2018, 1–13.
- [23] Claudino, J.G., Capanema, D.d.O., de Souza, T.V., Serrão, J.C., Pereira, A.C.M., Nassis, G.P., 2019. Current Approaches to the Use of Artificial Intelligence for Injury Risk Assessment and Performance Prediction in Team Sports: a Systematic Review. *Sports Medicine - Open* 5, 1–12
- [24] Kumar, A., Gandhi, J., 2019. A Survey on Sports Prediction using Machine Learning. *Journal of Applied Science and Computations* 6, 1419–1423.
- [25] Fox, J.L., Scanlan, A.T., Sargent, C., Stanton, R., 2019. A survey of player monitoring approaches and microsensor use in basketball. *Journal of Human Sport and Exercise* 15, 230–240.
- [26] Pernigoni, M., Conte, D., Calleja-González, J., Boccia, G., Romagnoli, M., Ferioli, D., 2022. The Application of Recovery Strategies in Basketball: A Worldwide Survey. *Frontiers in Physiology* 13, 1–7.
- [27] Guillén, S., Arredondo, M.T., Castellano, E., 2011. A Survey of Commercial Wearable Systems for Sport Application. Springer. Healy, L., Tincknell-Smith, A., Ntoumanis, N., 2018. *Goal Setting in Sport and Performance*. Oxford University Press, Oxford, England.
- [28] Ometov, A., Shubina, V., Klus, L., Skibińska, J., Saafi, S., Pascacio, P., Fluoratoru, L., Gaibor, D.Q., Chukhno, N., Chukhno, O., Ali, A., Channa, A., Svertoka, E., Qaim, W.B., Casanova-Marqués, R., Holcer, S., Torres-Sospedra, J., Casteleyn, S., Ruggeri, G., Araniti, G., Burget, R., Hosek, J., Lohan, E.S., 2021. A Survey on Wearable Technology: History, State-of-the-Art and Current Challenges. *Computer Networks*
- [29] Junior, J.E.H., Schack, T., 2017. Integrating Pre-Game rituals and PrePerformance routines in a culture-specific context: Implications for Sports Psychology Consultancy. *International Journal of Sports and Exercise Psychology* 17, 1–14.
- [30] Todd, M., Brown, C., 2003. Characteristics Associated with Superstitious Behaviour in Track and Field Athletes: Are there NCAA Division level Differences. *Journal of Sports Behaviour* 26, 128–187.

- [31] Raalte, J.L.V., Brewer, B.W., Nemeroff, C.J., Linder, D.E., 1991. Chance orientation and superstitious behavior on the putting green. *Journal of Sport Behavior* 14, 41–50.
- [32] Prapavessis, H., 2000. The POMS and sports performance: A review. *Journal of Applied Sports Psychology* 12, 34–48.
- [33] Mazlan, R., Saim, L., Thomas, A., Said, R., Liyab, B., 2002. Ear Infection and Hearing Loss Amongst Headphone Users. *The Malaysian Journal of Medical Sciences* 9, 17–22.
- [34] Terry, P.C., Karageorghis, C., Curran, M., Martin, L., Parsons-Smith, R., 2020. Effects of Music in Exercise and Sport: A Meta-Analytic Review. *Psychological Bulletin* 146.
- [35] Jones, J.W., Neuman, G., Altmann, R., Dreschler, B., 2001. Development of the Sports Performance Inventory: A Psychological Measure of Athletic Potential. *Journal of Business and Psychology* 15, 491–503.
- [36] Cunha F, Heckman J, Schennach S. Estimating the Technology of Cognitive and Noncognitive Skill Formation. *Econometrica*. 2010 May 1;78(3):883-931.
- [37] Waldron, J.J., Krane, V., 2005. Whatever it Takes: Health Compromising Behaviors in Female Athletes. *Quest* 57, 315–329.
- [38] Fox, T.P., Oliver, G., Elis, S.M., 2013. The Destructive Capacity of Drug Abuse: An Overview Exploring the Harmful Potential of Drug Abuse Both to the Individual and to Society. *International Scholarly Research Notices* 2013, 1–6.
- [39] Gable, R.S., 1993. Toward a Comparative Overview of Dependence Potential and Acute Toxicity of Psychoactive Substances Used Nonmedically. *The American Journal of Drug and Alcohol Abuse* 19, 263–281.
- [40] Gable, R.S., 2004. Comparison of acute lethal toxicity of commonly abused psychoactive substances. *Addiction* 99, 686–696.
- [41] Evins, A.E., Green, A.I., Kane, J.M., Murray, R.M., 2013. Does using marijuana increase the risk for developing schizophrenia? *Journal of Clinical Psychiatry* 74, e08
- [42] Moses, E., Dunn, P., Smith, T., Clarke, J.B., Wright, K., Davis, T., Matsumoto, A.M., Merrens, E.J., Plummer, D., Rosen, P.M., . World AntiDoping Agency's (WADA) Prohibited List. Internet. URL: <https://www.usada.org/athletes/substances/prohibited-list/>. retrieved October 1st, 2000.
- [43] Reardon, C.L., Creado, S., 2014. Drug abuse in athletes. *Substance Abuse and Rehabilitation* 5, 95–105.
- [44] Taylor, K.L., Chapman, W.D., Cronin, J., 2012. Fatigue Monitoring in High Performance Sport: A Survey of Current Trends. *Journal of Australian Strength and Conditioning* 20, 12–23.
- [45] Takanashi, Y., Fujimori, N., Koikawa, N., 2020. An investigation into the relationship between throw performance and maximum weight in weight training of female discus throwers. *Journal of Human Sport and Exercise* 16, 226–234
- [46] Bideau, B., Kulpa, R., Vignais, N., Brault, S., Multon, F., Craig, C., 2010. Using Virtual Reality to Analyze Sports Performance. *IEEE Computer Graphics and Applications* 30, 14–21.

# An Improvement for Spatial-Temporal Queries of ATMGRAPH

ZHANG Zhiyuan<sup>1</sup>, HAN Boyang<sup>2</sup>

School of Computer Science and Technology, Civil Aviation University of China, Tianjin, China, 300300

**Abstract**—As a knowledge graph for the field of ATM (Air Traffic Management), ATMGRAPH integrates aviation information from various sources, and provides a new way to comprehensively analyze ATM data, but the storage schema of ATMGRAPH is inefficient for trajectory-related queries which have typical spatial-temporal characteristics, thus cannot meet the application requirements. This paper presents an improved storage model of ATMGRAPH, specifically, we design a cluster structure to connect trajectory points and spatial-temporal information to speed up trajectory-related queries, and we link flights, airports, and weather information in an effective way to speed up weather-related queries. We create a dataset of about 10,000 real domestic flights, and build a knowledge graph of it which contains about 11.66 million triplets. Experimental results show that ATM knowledge graph constructed by this storage model can significantly improve the efficiency of spatial-temporal related queries.

**Keywords**—Air traffic management; knowledge graph; storage model; spatial-temporal query; ontology

## I. INTRODUCTION

With the rapid development of the economy, people are willing to travel by air due to its efficiency and convenience. The civil aviation industry generates a large amount of data every day, coming from multiple departments such as airports, airlines, Air Traffic Managements (ATMs) and meteorological bureaus, with varying data forms and coding rules, make it a great challenge for semantic data query and analysis. As Aviation data are scattered in different systems, integrating them into a big semantic database seems to be a good idea. The most representative work is ATMGRAPH (Air Traffic Management Knowledge Graph) constructed by NASA. This KG (Knowledge Graph) integrates multiple aviation information and is benefit for semantic data analysis. Flight trajectory information accounts for the vast majority in ATMGRAPH, it has obvious spatial-temporal characteristics, and data analysis on trajectory is often about spatial and temporal. However, in practical applications, ATMGRAPH encounters great scale problems, especially when facing spatial-temporal related data queries, i.e. its performance decreases dramatically for huge data volumes.

There are few works to address this problem, in order to fill this research gap, this paper conducts on spatial-temporal query optimization of ATMGRAPH. A knowledge graph can be logically divided into two layers: the data layer and the schema layer. The data layer stores knowledge facts, and the schema layer defines ontology to standardize a series of fact expressions in the data layer [7]. This paper designs an improved storage model for ATMGRAPH to solve the

problem of slow and inefficient processing of spatial-temporal related queries. Specifically, we design a cluster structure to connect trajectory points and spatial-temporal information to speed up trajectory-related queries, and we link flights, airports, and weather information in an effective way to speed up weather-related queries. Experimental results on real aviation data show that the query efficiency using our model is significantly improved in typical application scenarios.

The rest of this paper is organized as follows. Section II is the related work. Section III is the problem definition, which introduces NASA's original ATMGRAPH model and analyzes its shortcomings in spatial-temporal related queries. In Section IV, we introduce our improved ATMGRAPH model in detail. Section V is the experimental results and discussion, and we conclude our work in the final section.

## II. RELATED WORK

With the rapid development of the global transportation industry, air traffic flow has significantly increased. There were lots of research works on air traffic management such as airspace saturation, flight accidents, flight delays, and air control difficulties. The Federal Aviation Administration (FAA) used big data analysis to identify operational patterns, which can support the identification and prediction of airport data [2]. Rezo [3] introduced a paradox in aviation data processing and proposed a probable solution. Dorota [4] discussed the requirements of aviation data in Polish regulations and gave a practical proposal. Keller et al. [5] introduced a system for combining heterogeneous air traffic management with semantic integration techniques, which transforms data from disparate source formats into a unified semantic representation of ontology-based triplets. Liu et al. [6] implemented seamless communication and mutual cooperation between civil aviation systems through information sharing, which could support collaborative decision-making of air traffic management and improve the capacity of airspace systems. Lu et al. [7] proposed an integration architecture of cloud computing and blockchain for ATM systems, in which it pointed out the advantages of the new technology architecture over the traditional architecture of existing ATM systems. Europe and the United States are trying to use ontology technology to integrate and fuse aviation data from multiple sources, so as to provide a unified data exchange mechanism with semantic information for all participants in the aviation industry. For example, the Single European Sky Program launched the BEST project (<http://www.project-best.eu>), which designed AIRM (ATM Information Reference Model) and constructed an ontology



model for aeronautical and meteorological information [8]. At the same time, NASA constructed ATMONGO (ATM Ontology), involving ATM core data such as aircraft, flight, airport, airline, route, and navigation facility [9]. It includes over 150 classes, over 150 datatype properties, and over 100 object properties. Based on ATMONGO, NASA also built ATMGRAPH, a knowledge graph containing 260 million triplets [10]. Many information of ATM has temporal and spatial characteristics, e.g. when an airport is temporarily closed due to snow conditions, the airport operation status in KG should be changed to CLOSED, and the start and end time should also be indicated. Therefore, Schuetz et al. [11] proposed the concept of Contextualized Knowledge Graphs by adding semantic dimensions such as time, space, and data source in KG to solve the problem of information distribution and acquisition for all participants in the aviation industry.

### III. PROBLEM DEFINITION

As the latest achievement of symbolism, knowledge graph is an important milestone of artificial intelligence. Knowledge graph can provide valuable structured information by data integration and standardization, and it has been widely used in information retrieval, automatic question answering, decision making and other fields, and it is also an important basic technology to promote data mining and intelligent information services [12]. With the growing scale of the knowledge graph, data management issues become increasingly prominent [13]. KG is generally divided into general knowledge graph and domain knowledge graph, and the latter usually needs to carefully design the storage model according to the industry data's characteristics in order to meet the retrieval requirements under large-scale data.

Consider the following two representative queries in ATM:

- Find all flights passing through the ZBAAAR20 sector of Beijing on July 20, 2022 and landing at Beijing Capital International Airport under strong wind conditions.
- Find which sector controlled the most flights between 9am and 10am on July 16, 2022.

ATMGRAPH consists of one month's flights (approximately 100,000 flights) and weather data in the New York metropolitan area, which includes eight classes: airspace structure and facilities, flight routes and procedures about takeoff and landing, traffic management measures, flight carriers and aircrafts related, airport and ground operations, weather, sequence related, and spatial-temporal related. Fig. 1 is a segment of ATMGRAPH, with a specific flight instance at its center: UAL535, which took off at 00:19:00 on July 15, 2014. Connected to it includes the departure and arrival airport of the flight, the carrier airline, the aircraft model, the planned route and the actual route. The lower part of the figure represents the track points of the flight, each contains information such as time, longitude, latitude, altitude, speed etc. (not listed in the figure for brevity). Although there are classes about weather and sectors in ATMGRAPH, getting the results of the two representative queries above is very inefficient, cause it must check all points

one by one whether it matches the corresponding constraint. For example, when querying the workload of a sector during a certain period of time (i.e. the number of flights flying within the sector during this time), at first we must find all track points within that period of time, then for all of them we need to check whether their positions are within that sector, and finally output the corresponding flight information. Obviously, these kinds of operations are quite inefficient.

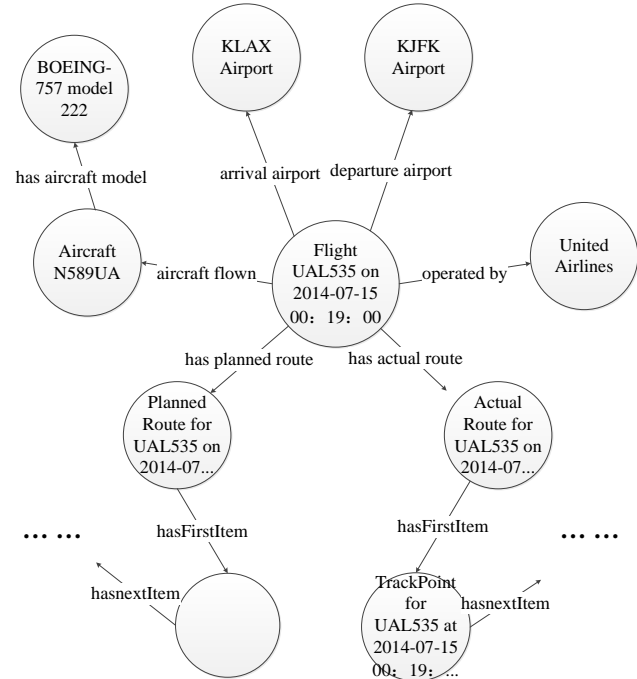


Fig. 1. Flight information storage segment.

This paper designs an improved storage model, which not only considers the strongly correlated characteristics between flight and weather information, flight and spatial-temporal information, but also links the trajectory points with spatial-temporal entities, so as to speed up spatial-temporal related queries. Experimental results on real flight data show that our proposed model greatly improves the query speed for representative queries and for some queries which the original model may take hours, our model can finish them in just a few seconds.

### IV. IMPROVED STORAGE MODEL

Although there are already eight major classes in ATMGRAPH to represent various knowledge in ATM field, some of them are relatively independent, making it difficult to obtain results using a single query statement involving multiple classes. The nodes of flight trajectories in ATMGRAPH account for nearly 70% of the entire graph, and each track point is only connected in chronological order using the hasNextTrackPoint relationship. This kind of storage model not only occupies a large amount of storage space but also reduces query efficiency. On the premise of being consistent with the original structure of the ATMGRAPH, this paper extends it to express more spatial-temporal information without taking up more storage space.

Fig. 2 illustrates our improved storage model, where ellipses represent the newly added classes and dashed edges represent the newly added relationships. TimeInterval is a new class for standard time segment, which connects the Trackpoint class through belongToTimeInterval relationship to express track points with the standard time segment information. The relationship belongToSector connects the Trackpoint class with the Sector class representing which sector the track point is located in. The new class WeatherInterval represents weather conditions of each airport in different time periods, and it also connects to the Flight class through two new edges: hasArriveWeather and hasDepartureWeather. The class ActualRoute represents the actual flight route, which contains the first and last track points of the trajectory through the edge of hasFirstTrackpoint and hasLastTrackpoint.

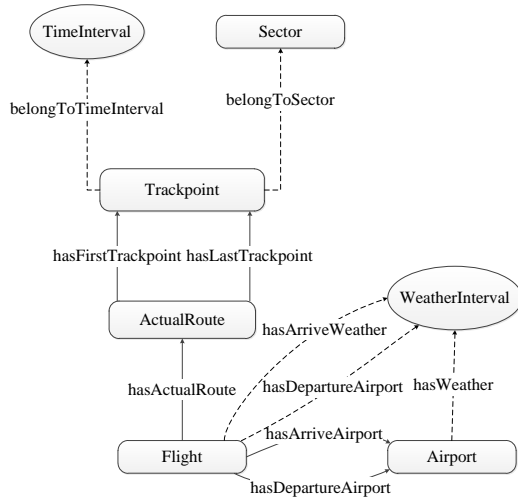


Fig. 2. Improved storage model of ATMGRAPH.

Through this storage model, track points has a direct connection to time and sector information, thus alleviates the problem of inefficient spatial-temporal related queries in the original ATMGRAPH. At the same time, weather information also has a direct connection to flights, which can solve the problem of slow query speed for weather and flight related queries.

#### A. Standard Time Interval

In the real world, many facts have time attributes, which play an important role in knowledge graph [14]. For example, the fact represented by a triplet (Steve Jobs, diedIn, California) is that Steve Jobs died in California, which occurred on October 5, 2011; The fact (Ronaldo, playing for A.C. Milan) was only valid between 2007 and 2008. In air transportation, when an aircraft performs a complete flight mission, time

information cannot be ignored. In our experiment, the flight data comes from ADS-B, which broadcasts real-time information including aircraft position, speed, identification code, flight number, and air-ground status to ATC (Air Traffic Control system) or other aircrafts through the air-to-air and air-to-ground data links [15]. Table I shows an ADS-B data fragment that contains three track points of flight EPA6206 during its mission on July 27, 2020. The specific information includes the flight number, aircraft number, and the current position (longitude, latitude, altitude), speed, heading, and data transmission time expressed in UTC (Universal Time Coordinated).

In ATM data analysis, usually we do not care much about the instantaneous state of an aircraft at a specific time point, and the time unit in queries is mostly hours or days. For example, finding the number of flights flying at altitudes above 6000m from 8:00 to 10:00 on July 10, 2022. To quickly retrieve the flight status of many flights within a same time segment, creating standard time intervals seems to be a feasible and effective method. This paper takes ten minutes as a standard time interval. If it is too long, it will lead to too many track points within a time interval, which will affect the query performance. If it is too short, it will cause too many TimeInterval nodes in the graph, and waste the storage space. Track points belonging to a same time interval are all linked to the TimeInterval entity representing that time segment, forming a cluster structure. Fig. 3 shows some track points of flight KNA8202 and flight CSZ9106 during the time interval from 7:00 to 7:10 on July 18, 2022. It can be seen that this cluster structure can gather all track points within the standard time interval without damaging the original relationships in the graph. Due to the fact that each track point is connected to its corresponding standard time interval node, related track points can be directly retrieved, without checking all track points one by one to judge whether they meet time constraints. This provides a more efficient way for time related query tasks.

The process of adding standard time intervals is as follows: Create all standard time intervals in the graph, and then calculate the corresponding TimeInterval for each track point according to UTC Time, and connect it with the relationship belongToTimeInterval. After that, the above query can be solved through a single Cypher query statement:

```

match(n:Trackpoint)-[r:belongToTimeInterval]-
(m:TimeInterval)
where n.height >= 6000
and m.startTime >= 2022/07/10 08:00:00
and m.endTime <= 2022/07/10 10:00:00
    
```

TABLE I. ADS-B DATA SEGMENT

Fnum	UTC Time	Latitude	Anum	Angle	Speed	Height	Longitude
EPA6206	2022/7/27 11:59:16	30.5709	B204N	21	361.14	1013.46	103.94346
EPA6206	2022/7/27 11:59:31	30.61871	B204N	22	355.584	1226.82	103.96566
EPA6206	2022/7/27 11:59:46	30.63263	B204N	22	357.436	1325.88	103.97208

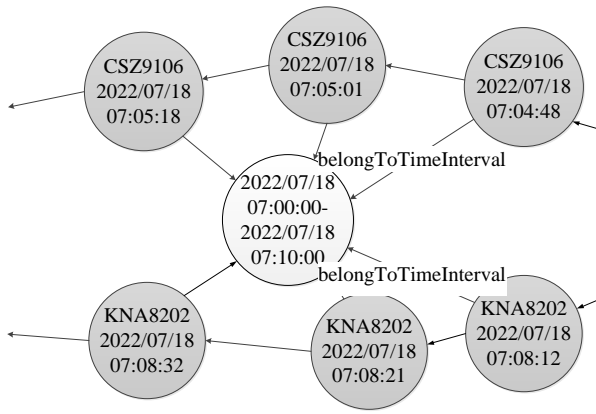


Fig. 3. Temporal cluster structure based on TimeInterval.

### B. Spatial Sector Clusters

Spatial is an essential attribute of geographic data, which is mainly used to describe the spatial features of geographical entities, including position, shape, and spatial relationships [16]. Sector is a major geospatial entity in ATMGRAPH, usually formed as a polygon with height range, and the polygon is consisted of multiple points with longitude and latitude coordinates connected from head to tail. Sector is a fundamental unit of air traffic management services and is an important component for airspace planning and allocation [17]. In the field of ATM, many typical queries focus on the workload of a sector over a period of time (i.e. the number of flights in that sector within a specific time period). For example, finding the workload of sector ZBAAAR18 from 8:00 to 10:00 on July 10, 2022.

Traditional way to answer this kind of question in ATMGRAPH is to check trajectory points one by one if it is located in the given sector, which is very time consuming. To address this, this paper takes the sector as a central node and connects all track points within the sector to it, thus forming a cluster structure. When executing above queries, we only need to search the corresponding sector first, and find all track points connected to it. After filtering out duplicate flight numbers, the query results can be obtained immediately. Due to the fact that adjacent trajectory points may belong to two different sectors, how to determine whether a track point is within a sector? The ray crossing number method is generally used to determine whether a point falls inside a polygon. Specifically, firstly we draw a ray emitted from that point, and then we calculate the number of intersections between the ray and the polygon boundary: if the number of intersections is odd, then the point is inside the polygon, otherwise it is outside the polygon. In Fig. 4, a ray passes through an irregular polygon, and if the starting point of the ray is located in the thin line section, it has an even number of intersections with the polygon, and if it is located in the thick line section, it has an odd number. According to the above method, the points in the thick line section are inside the polygon, and the points in the thin line section are outside.

Fig. 5 shows a cluster structure fragment centered on sector ZBAAAR18 in the Beijing flight control area, with surrounding nodes of trajectory points. The names displayed in the nodes are the flight numbers and instantaneous times.

Similar to the temporal cluster, this structure also does not disrupt the original track point connection relationship in the graph. When conducting a query about workload of an air traffic control sector, it is possible to directly find the Sector node and use the belongToSector relationship to reversely find its connected track points, and it is not necessary to calculate the position of each track point any more, thus greatly reducing the query time.

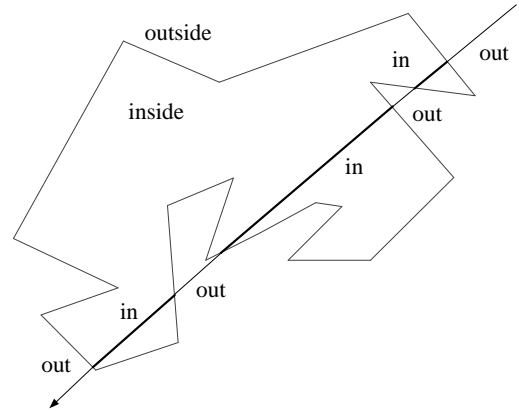


Fig. 4. Ray crossing number method to determine whether a point is inside the polygon.

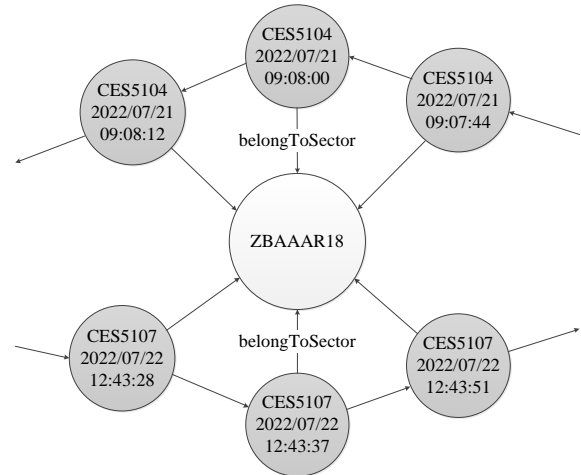


Fig. 5. Spatial cluster structure based on sector.

The process of adding spatial sector nodes is as follows: First, import the information of all sectors into the knowledge graph, then use ray crossing number method to calculate each track point to judge its relationship to the sectors, and finally connect each track point with its sector through the belongToSector relationship. With the spatial cluster structure, it is very easy to answer sector related queries. For example, when solving the query mentioned in this section, we can obtain the results in Neo4j by a single Cypher query statement:

```
match(n:Trackpoint) -[r1:belongToTimeInterval]-
(m:TimeInterval)
```

where  $m.startTime \geq 2022/07/10\ 08:00:00$

and  $m.endTime \leq 2022/07/10\ 10:00:00$

```
with n match(n)-[r2:belongsToSector]-
(o:sector{name:'ZBAAAR18'})
return count(distinct(n.fnum))
```

After adding temporal and spatial clusters to ATMGRAPH, all track points are connected with their corresponding temporal and spatial information. Fig. 6 shows a segment about the connection relationship between TimeInterval, sector and Trackpoint of our improved knowledge graph. At this point, when considering a query like 'Which sector controlled the most flights between 9:00 am and 10:00 am on July 16, 2022', we can simply find the six TimeIntervals that represent this period of time, filter out all the track points within them, and then count the number of flights included in each sector to obtain the final results.

### C. Airport Nodes with WeatherIntervals

Weather conditions are very important for aircraft takeoff and landing. For the first representative query mentioned in Section II about finding all flights that pass through the ZBAAAR20 sector and land at Beijing Capital International Airport (BCIA) under strong wind conditions on July 20, 2022, the usual processing method requires two query operations (querying the time span of strong wind conditions at BCIA that day, and querying all flights that land at BCIA that day) and one comparison operation (check those flights one by one if its landing time is in the time period of strong wind). If the weather information when an aircraft arrives at or departs from an airport is directly stored in the knowledge graph, then the speed of answering such questions will be significantly improved.

This paper adds weather information to each airport based on the class WeatherInterval, and each flight is also directly connected to its weather information during departure and arrival. Considering that weather generally does not change frequently in a short period of time, unlike TimeInterval, the span of the WeatherInterval is set to 12 hours. As shown in Fig. 7, flight CHH7810 landed at Beijing Capital International Airport on July 18, 2022, and the weather when landing was rainy. Airports may have different weather conditions at different time periods and are connected to WeatherInterval by the relationship of hasWeather. Flights are also connected to WeatherInterval by relationships of hasArriveWeather and hasDepartureWeather. Using the new schema, when processing queries related to weather conditions, there is no need to match the landing time and corresponding weather information of flights one by one anymore, and it can be obtained directly through WeatherInterval. For the typical query mentioned in this section, we can obtain the results in Neo4j by a single Cypher query statement:

```
match(n:flight)-[r:arriveAirport]-(m:Airport{code:'PEK'})
where n.endtime >= 2022/07/20 00:00:00
and n.endtime < 2022/07/21 00:00:00
with n match(n)-
[:hasArriveWeather]- (:weatherInterval{weather: 'strong
wind'})
```

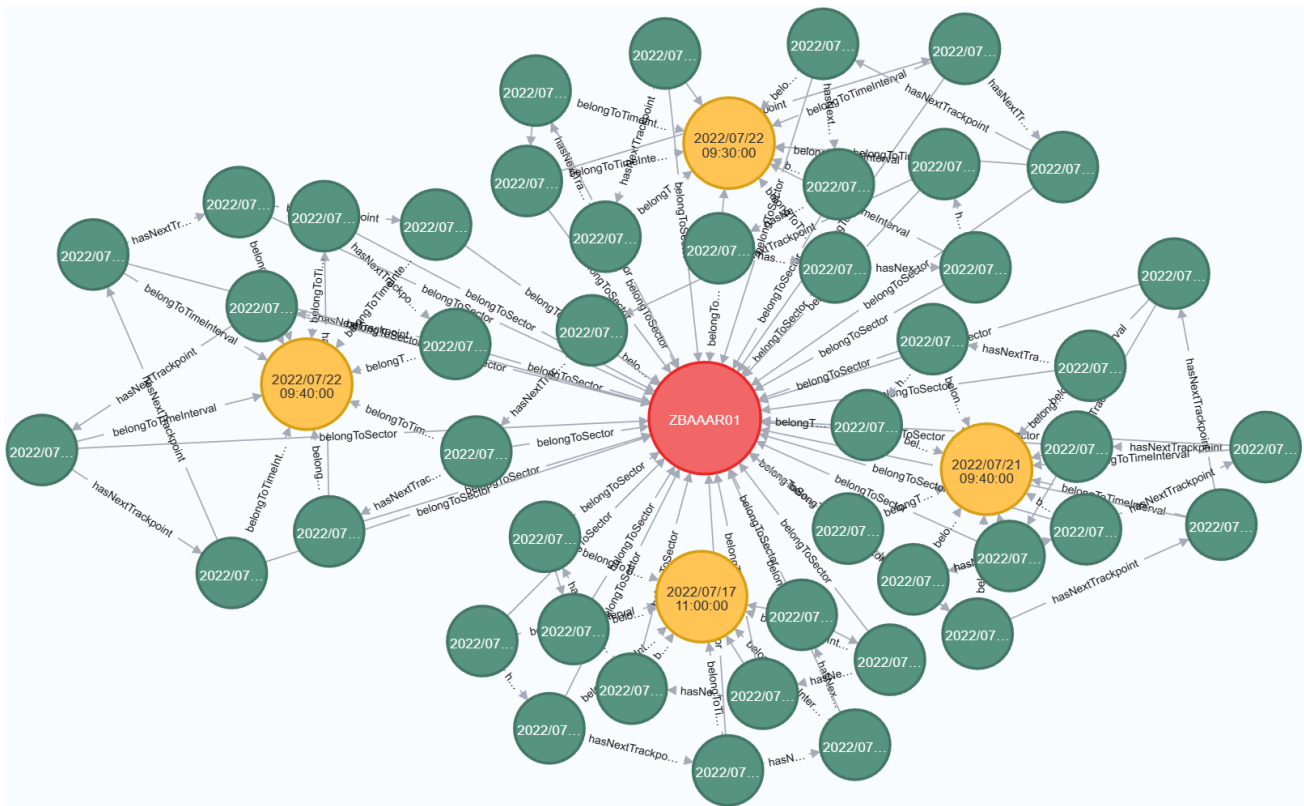


Fig. 6. A fragment of our improved ATMGRAPH.

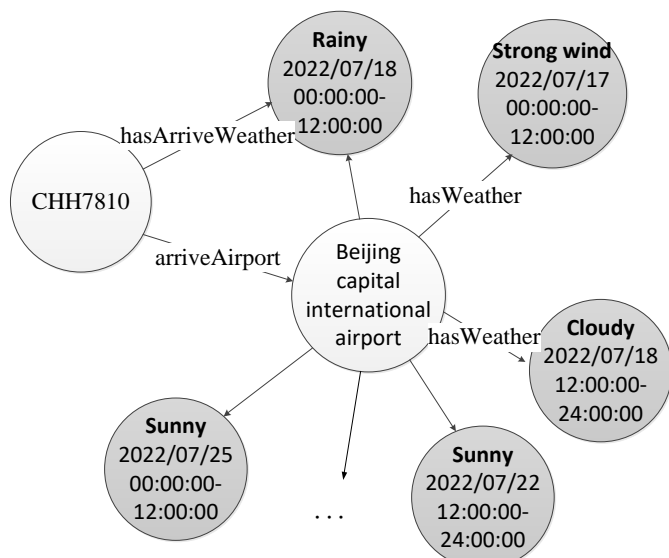


Fig. 7. A fragment of flights, airports and WeatherIntervals.

### V. EXPERIMENTAL RESULTS AND DISCUSSION

The experimental environment is a 64bit Windows system (Intel i7-7700HQ CPU, 16GB memory), 4.2.2 community version Neo4j, implemented using Python language.

We crawled ADS-B data from varflight website (<https://www.variflight.com>) about 10,000 flights from July 16 to 27, 2020. The data of airports, flight information regions, and sectors are from AIP (Aeronautical Information Publication), and weather data is randomly set. Using these data, we built ATMGRAPH of two versions: NASA's original version and our improved version, with the latter containing 5.5 million nodes and 11.66 million relationship edges. We then evaluate their performance using three typical query cases, and the results are shown in Table II. The evaluation metric is query time. The first query case is only temporal related, the second query case is spatial-temporal related, and the third query case is about time, airport, and weather condition.

Table II shows the comparison results between NASA's original ATMGRAPH and our improved version on commonly used spatial-temporal related queries. The first is a common time related query in ATM data analysis. Using our storage model, due to the existence of standard time segment clusters, it is very easy to find all track points belonging to the TimeInterval from 9:00 to 12:00 on July 27, 2020. On the contrary, for the original ATMGRAPH we must compare the UTC value in each track point. From the results in Table II, we can see that our model is about eight times faster than ATMGRAPH. The second query adds spatial constraint to the first one. For ATMGRAPH, because there is no direct connection between track points and flight information regions, to get the query results, we must calculate all track points in the graph and judge the topological relationship between each track point and each flight sector. Because the number of track points is very huge and grows lineally with flight numbers, plus the position calculation is also very complex, thus it takes hours to obtain the query result. After adding a spatial cluster structure in our improved model, track points in the specified region can be directly found through the relationship belongToSector, and then the corresponding number of flights can be quickly obtained. The third query is related to the weather conditions at landing time. For this query, our model can directly get the flights that meet the conditions through a simple Cypher statement, while the original ATMGRAPH can be very complex: it should first identify the flights that land at the airport, and then find weather information of the airport during the landing time of the flights. Due to the fact that weather and flights in the original ATMGRAPH are not connected, the analyzer must manually check these flights one by one which is very time consuming, or develop a program to handle it which is very inconvenient. The result of ATMGRAPH for the 3rd query in Table II is gotten in a program way, which is about two seconds, while our model only uses five milliseconds, more than 400 times faster.

The above experimental results and discussion indicate that adding the spatial-temporal cluster structure proposed in this paper to ATMGRAPH can quickly process queries related to spatial-temporal features and improve data analysis speed.

TABLE II. PERFORMANCE COMPARISON OF TYPICAL QUERIES

Query cases	ATMGRAPH Version	Query Time
Find the number of flights from 9:00 to 12:00 on July 27, 2022	ORIGINAL	3748ms
	IMPROVED	421ms
Find the number of flights passing over Beijing on July 25, 2022	ORIGINAL	2.5h
	IMPROVED	1346ms
Find the number of flights landed at Beijing Capital International Airport from July 16 to 27, 2022	ORIGINAL	2160ms
	IMPROVED	5ms

## VI. CONCLUSION

In order to solve the problem of low efficiency of spatial-temporal related queries in ATMGRAPH, this paper proposes an improved storage model, which uses spatial-temporal clusters to represent flight information regarding time and location. In our improved model, trajectory points are connected to standard time intervals and sectors, and flight and airport entities are connected to weather intervals. Experimental results show that after adding the spatial-temporal cluster structure to the knowledge graph, the speed of relevant queries is greatly improved.

Due to the fact that the track data in ATMGRAPH accounts for approximately 70% of the total data volume, this article only focuses on improving the mode layer and cannot solve the redundancy problem of a large amount of track data. Aircraft trajectory points are stored as an unidirectional chain structure in Neo4j, and we can study a new storage structure for this typical kind of data in the future to save storage space and to optimize data query speed.

## REFERENCES

- [1] Z. Xu, Y. Sheng, L. He, and Y. Wang, Review on knowledge graph techniques[J]. Journal of University of Electronic Science and Technology of China, 2016, 45(04): 589-606. (in Chinese)
- [2] FAA. Terminal area forecast (TAF) [EB/OL]. [2017-05-18]. <https://taf.faa.gov>
- [3] R. Zvonimir, M. Tomislav, S. Sanja, and T. Andrea, A Paradox in aeronautical data processing: A case study review[J]. Case Studies on Transport Policy, 2022, 10(2).
- [4] D. Marjańska, Aeronautical data requirements and geodetic data – a case study on regulations in Poland[J]. Aircraft Engineering and Aerospace Technology, 2022, 94(5).
- [5] R. M. Keller, S. Ranjan, M. Wei, and M. M. Eshow, “Semantic representation and scale-up of integrated air traffic management data,” SBD '16, 2016.
- [6] L. Liu, H. Yang, and Y. Huang, Sharing big data storage for air traffic management[C]. 2022 IEEE 8th International Conference on Computer and Communications (ICCC), Chengdu, China, 2022, 1199-1203, DOI: 10.1109/ICCC56324.2022.10065656.
- [7] X. Lu, and Z. Wu, “ATMCC: Design of the Integration Architecture of Cloud Computing and Blockchain for Air Traffic Management.” 2021 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCLOUD/SocialCom/SustainCom) (2021): 37-43.
- [8] I. Kovacic, D. Steiner, C. Schuetz, B. Neumayr, and S. Wilson, Ontology-based data description and discovery in a SWIM environment[J], ICNS 2017, 1-22, doi: 10.1109/ICNSURV.2017.8012006.
- [9] R. M. Keller, The NASA air traffic management ontology: Technical Documentation Technical Memo NASA/TM-2017-219526, National Aeronautics and Space Administration, <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20170006095.pdf>, 2017.
- [10] R. M. Keller, Building a knowledge graph for the air traffic management community[C]// Companion The 2019 World Wide Web Conference. 2019.
- [11] E. Gringinger, B. Neumayr, S. Christoph, M. Schrefl, and S. Wilson, The case for contextualized knowledge graphs in air traffic management[C]//CKG SemStats@ ISWC. 2018.
- [12] T. Hang, J. Feng, and J. Lu, Knowledge graph construction techniques: Taxonomy, survey and future directions[J]. Computer Science, 2021, 48(02): 175-189. (in Chinese)
- [13] X. Wang, L. Zou, C. Wang, P. Peng, and Z. Feng, Research on knowledge graph data management: A Survey[J]. Journal of Software, 2019, 30(07): 2139-2174. DOI: 10.13328/j.cnki.jos.005841. (in Chinese)
- [14] T. Jiang, T. Liu, T. Ge, L. Sha, B. Chang, and S. Li, Towards time-aware knowledge graph completion[C]//Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers. 2016: 1715-1724.
- [15] K. Li. Research on the application status and trends of ADS-B abroad[J]. Aerodynamic Missiles Journal, 2018, 1(12): 60-66. (in Chinese)
- [16] J. Liu, H. Liu, X. Chen, X. Guo, and X. Zhu, Construction of knowledge graph based on Geo-Spatial data[J]. Journal of Chinese Information Processing, 2020, 34(11): 29-36. (in Chinese)
- [17] C. Xu, Y. Tian, K. Niu, W. Gong, and G. Li, Review of optimization for airspace sectorization[J]. Aeronautical Computing Technique, 2022, 52(01): 126-130. (in Chinese)

# Comparison of Machine Learning Algorithms for Crime Prediction in Dubai

Shaikha Khamis AlAbdouli<sup>1</sup>, Ahmad Falah Alomosh<sup>2</sup>, Ali Bou Nassif<sup>3</sup>, Qassim Nasir<sup>4</sup>

Dept. of Sociology, University of Sharjah, Sharjah, UAE<sup>1,2</sup>

Dept. of Computer Engineering, University of Sharjah, Sharjah, UAE<sup>3</sup>

Dept. of Computing and Informatics, University of Sharjah, Sharjah, UAE<sup>4</sup>

**Abstract**—This study aims to find the most accurate algorithm that is capable of predicting crimes in Dubai. It compares models on a dataset of sample crimes in the Emirate of Dubai, United Arab Emirates using the open-source data mining software WEKA, which enabled us to use Random Forest, KNN, SVM, ANN, Naïve Bayes and Decision Tree. We chose those algorithms as former studies that were effective used them. We have applied the algorithms on a dataset containing 13440 Major Crime in four categories occurred between 2014 and 2018. After comparing the models and analyzing their success rates, we identified the ideal algorithms and evaluated the effectiveness of variables in making predictions by measuring the correlation coefficients. One of the study's most crucial recommendations is to increase the variables and data, also adding more details about the crime, the criminal, and the victim. These variables make an impact on the analysis and the ultimate prediction.

**Keywords**—Machine learning; crime analysis; crime patterns; KNN; random forest; SVM; ANN; Naïve Bayes; Decision Tree; major crime

## I. INTRODUCTION

The modern society in Dubai has gathered people from around all the world, more than 170 nationalities, estimated at 3,478,300 people in 2021 [1], Dubai is well known for the low crime rate [2], in Q3 of 2022, Dubai Police has reported 65% drop in the number of criminal reports at the General Department of Criminal Investigations (CID) quarterly appraisal meeting, which was presided over by Lieutenant General Abdullah Khalifa Al Marri, Commander-in-Chief of Dubai Police [3].

The rise in urban crime statistics has become a major concern for law enforcement agencies across the world. Machine learning algorithms have been progressively used to predict and prevent crime in recent years. We intend to compare the performance of various machine learning algorithms for crime prediction in Dubai in this applied scientific research. Crime is a pervasive, global social problem that lowers people's quality of life and slows economic growth [4]. As it affects people's security, crime reduction remains one of the most important social issues in large metropolitan areas [5].

In order to reduce crime by predicting and preventing it, we must have a clear understanding of the current crime situation, which requires a crime data set that enables the use of machine learning. Predicting the future occurrence of crime is more possible today than ever before with digitalization and e-

governance generating data that allows for effective analysis [6].

We hope to gain insights into the most effective methods for predicting and preventing crime in this city by analyzing and comparing the accuracy, precision, and recall of these algorithms. Some research claims that crime cannot be predicted, as predictions are never 100% accurate [7]. Indeed, data is not always helpful in solving real world problems, but some scholars have succeeded in building models that helped to prevent crime [8]. This suggests that the issue with prediction may sometimes be caused by using the wrong model. Predictive policing aims to identify areas that may be subject to crimes. This is supported by routine activity theory and rational choice theory. According to both theories, a crime occurs when a person who is willing to commit it has the chance to do so and these opportunities follow patterns in both location and time rather than being distributed randomly [9].

This paper is structured as follows: Section II presents the main problem and motivation for the work. Section III presents work related to this research. Section IV describes the methodology. Section V presents the prediction models we use to analyze the data. Section VI presents our results, and Section VII summarizes our conclusions and related work.

## II. PROBLEM AND MOTIVATION

There are no applied, academic studies open to students based on Dubai Crime Data as they due to the restrictions which keeps access to crime data internal and confidential.

With the unprecedented support of the Dubai Police, this applied study gave us access to real crime data in Emirates of Dubai.

By this research, we are trying to find the most accurate algorithm that is capable of predicting crimes in Dubai.

## III. RELATED WORK

There are very few similar studies in the Arab Region so far. Scholars tend to conduct theoretical research and surveys, and not real crime data-based studies, On the other hand, we have found countless examples of work from other regions.

- Crime Rate Prediction Using Machine Learning and Data Mining by Sattar, Abdus and others [10] uses different clustering approaches of data mining to analyze the crime rate of Bangladesh. The authors use the KNN algorithm, and identify geographical areas that

have higher crime rates, making recommendations for individuals to be cautious in those areas.

- Crime Analysis and Prediction Using Machine Learning by Olta Llaha [11] identifies the most appropriate data mining methods for analysing data collected from crime prevention sources by theoretically and practically comparing them. The authors use gender, age, employment status, and crime location as attributes. They find that data mining methods help to predict the incidence of a crime occurring and, as a result, contribute to avoiding it.
- An Experimental Study of Crime Prediction Using Machine Learning Algorithms by Sikhnam Nagamani and others [12] uses open data from Kaggle, a mix of crime types, description, time and date, and latitude and longitude to find patterns in crimes.
- Comparison of Machine Learning Algorithms for Predicting Crime Hotspots by XU ZHANG and others [13] uses an open data source from China (2015 to 2018). It suggests the use of historical crime data as well as covariates associated with criminological theories in order to evaluate the merit of machine learning algorithms.
- Crime Prediction through Urban Metrics and Statistical Learning by Luiz G. A. Alves and others [14] uses random forest regressor to predict crime and quantify the influence of urban indicators on homicides. This study finds that random forest algorithm is an excellent model for predicting crime.
- Using Machine Learning Algorithms to Analyze Crime Data by Lawrence McClendon and Natarajan Meghanathan [15] uses WEKA, open-source data mining, to conduct a comparative study between the violent crime patterns from the Communities and Crime Unnormalized Dataset provided by the University of California-Irvine repository, and actual crime statistical data for the state of Mississippi that has been provided by neighborhoodscout.com. This study finds the linear regression algorithm to be very effective and accurate in predicting crime data based on the training set input for the three algorithms.

The current study makes significant contributions by attempting to fill multiple research gaps.

First, the study adds to the relatively limited research on crime prediction in the Arab world. Our study is one of the first to use prediction models on real data from a reliable source in the Arab region.

Second, ours is one of the few research projects that has used six prediction models to determine which provides the best outcome with greater understanding and insight into the data used.

Third, to the best of the author's knowledge (based on a search of peer-reviewed databases), no previous study has compared machine learning algorithms on crime prediction in Dubai in an applied academic setting.

## IV. METHODOLOGY

### A. Dataset

The crime data used in this study is confidential data individually supplied to the research team by the Dubai Police. The only publicly available crime data in Dubai are the total published by the Dubai Police, which would be insufficient for the completion of the present study [16]. This restricted and non-georeferenced dataset consists of a spread sheet that contains data compiled by the police, as shown in Table I, containing the date, the hour, the typology, used tool, the technique used, and the area of the crime, as well as the age, nationality, status and education level of the criminal for all reported crimes occurring inside the city limits between January 2014 and December 2018, amounting to approximately 52 thousand entries.

TABLE I. DATA SET USED

Name	Description	Data Type
Date	Date of crime	Date
Time	Time of crime	Nominal T1: 12:00 am to 5:59am T2: 6:00am to 11:59 am T3: 12:00pm to 5:59pm T4: 6:00pm to 11:59pm
Police	Police Station responded to the crime, which refers to the area as well	String
Age	Age of criminal	Numerical
Sex	Sex of criminal	String
Nationality	Nationality of Criminal	String
Education	Educations of Criminal	String
Status	Status of Criminal	String

### B. Pearson Correlation Coefficient

Pearson correlation coefficient; descriptive statistic; indicates relationship (extent of linear correlation) between two continuous variables; the better comparable the data resulting from two different methods are (i.e. the closer the correlation is) the more the r value approaches the value 1, whereby 0 represents no correlation, -1 a perfect inverse correlation (negatively sloping line) and +1 a perfect positive correlation [34].

We calculated the correlation coefficient value in order to determine how strong the association between the factors.

The correlation coefficient can be understood as follows:

- There is absolutely no association when the correlation coefficient is 0. It implies that the variables have a fully unfavorable connection. There is no association if the correlation coefficient is zero.
- If the correlation coefficient is 1, a significant positive correlation is demonstrated. It implies that the variables have their optimal positive correlation.
- A correlation coefficient with a larger absolute value denotes a stronger link between the variables.



We applied Pearson correlation in the crime dataset using Weka by selecting the attribute ranking using correlation Attribute Eval. Here are the outcomes we obtained:

TABLE II. RANKED ATTRIBUTES

Ranked Attributes	
0.1055	Nationality
0.0999	Time
0.076	Date
0.0621	Status
0.0508	Sex
0.0495	Police
0.031	Education
0.025	Age

In the Table II, we notice the most significant attribute affect for crime type is nationality with a weight = 0.105. The second largest attribute is time with a weight = 0.099. The third largest attribute is date with a weight = 0.076. The fourth largest attribute is status with weight 0.062. Next is sex, with a weight of = 0.050. Then police, with a weight of = 0.049. Finally, the last two attributes are education with a weight of = 0.031, and age with a weight of = 0.025.

### C. Preprocessing

First, we selected four major crime typologies out of 10. The Major Crimes are categorized as: (Willful Murder, Aggravated Assault, Rape, Robbery, Theft, Abduction, Grand Auto Theft, Burglary, Drugs, Human Trafficking) due to the Non-disclosure Agreement we cannot declare which four categories we have chosen. We removed any crimes that had missing values due to missing data or compiling errors. This reduced the number of entries to 13,440.

Instead of using exact times, we categorized hours into four periods, 6am to 12pm, 12pm to 6pm, 6pm to 12am, and 12am to 6am. We categorized nationalities into three groups: Gulf Countries (Saudi Arabia, United Arab Emirates, Oman, Kuwait, Bahrain, Qatar), Arab countries (Algeria, Comoros, Djibouti, Egypt, Iraq, Jordan, Lebanon, Libya, Mauritania, Morocco, Palestine, Somalia, Sudan, Syria, Tunisia and Yemen), and rest of world.

To train and validate the data, the dataset is divided into various subsets with 10 folds in cross-validation, the training was on 70% and test was on 30%.

### D. Evaluation Metrics

1) *Accuracy*: The percentage of overall predictions that were correct.

2) *Accuracy*:  $((TP + TN) / (TP + FP + TN + FN)) * 100$

3) *Precision*: Precision reveals the proportion of genuinely positive forecasts among all positive ones. The ratio of accurately positive predictions to all positive predictions is how it is defined.

4) *Precision* = Predictions accurately positive / Total predicted.

5) *Precision* =  $TP / (TP + FP)$

6) *Recall*: Shows how many truly positive values were predicted out of all positive values. It measures the proportion of accurate positive predictions to all the positive examples found in the dataset.

7) *Recall*: It is the ratio of predicted values that came true to actual values in the dataset.

8) *Recall* =  $TP / (TP + FN)$

9) *F1 Score*: It is the harmonic mean for precision and recall values as depicted in Fig. 1. [17]

$$F_1 = \left( \frac{\text{recall}^{-1} + \text{precision}^{-1}}{2} \right)^{-1} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Fig. 1. F1 equation.

## V. PREDICTION MODELS

### A. Random Forest

During the training phase of the random forests or random decision forests ensemble learning approach (which is used for classification, regression, and other tasks), a large number of Decision Trees are built. For classification problems, the random forest output is the class that the majority of the trees choose. For regression tasks, the mean or average prediction of each individual tree is returned. Random decision forests correct Decision Trees' proclivity for overfitting their training dataset [18] [19] [20] [21].

The classifier used is Random Forest.

### B. (The K-Nearest Neighbor's Algorithm) KNN

The K-Nearest-Neighbours (KNN) is a non-parametric classification method, which is simple but effective in many cases [22]. KNN is a non-parametric classification algorithm, it works as a supervised learning algorithm. A labeled training dataset is provided where the data points are categorized into various classes, so that class of the unlabeled data can be predicted [23].

The classifier used in KNN is IBk.

### C. Support-Vector Machines (SVM)

Support vector machines (SVMs) can be used to handle classification, regression, and outlier problems that are frequently encountered in supervised learning [24].

The mathematical pedigree of SVMs is the best of any statistical learning procedure. It was created as a classifier that maximizes a slightly different definition of a margin, resulting in a novel "hinge" loss function [25]. Weka can classify objects using the support vector machines algorithm [26].

### D. Artificial Neural Networks (ANNs)

Artificial Neural Networks can be defined as systems designed to model functions that simulate the human brain [27]. They are increasingly being used to model complex, nonlinear phenomena [28]. ANNs are nonlinear, adaptive information processing systems that are made up of many interconnected processing units. ANNs have functions such as associative memory, nonlinear mapping, classification

recognition, and optimization computation as an effective empirical modelling tool [29].

The classifier used in ANN is Multilayer Perception.

### E. Naive Bayes

The naive Bayes classifier significantly simplify mastering through assuming that capabilities are impartial given class [30]. Naive Bayes is a probability classification model that makes machine learning easier by performing calculations on datasets with the goal of predicting probabilities in a class under the assumption of strong independence. Classification is a type of directed learning [31].

The classifier used in Naive Bayes is Naive Bayes

### F. Decision Tree

Decision tree is one of the popular predictive modelling approaches used in many areas including statistics, data mining and machine learning [32]. Decision tree classifiers are regarded to be a standout of the most well-known methods to data classification representation of classifiers [33].

The classifier used in Decision Tree is J48 which is a Decision Tree classification algorithm based on Iterative Dichotomiser 3.

## VI. RESULTS

We can summarize the results in Table III:

TABLE III. ALGORITHMS RESULTS

Algorithm	Accuracy	Recall	Precision	F1
Random forest	76.986%	0.77	0.77	0.769
KNN	78.474%	0.785	0.789	0.784
SVM	54.575%	0.546	0.524	0.520
ANN	51.093%	0.511	0.511	0.509
Naïve Bayes	53.526%	0.535	0.516	0.511
Decision Tree (J48)	67.976%	0.680	0.678	0.678

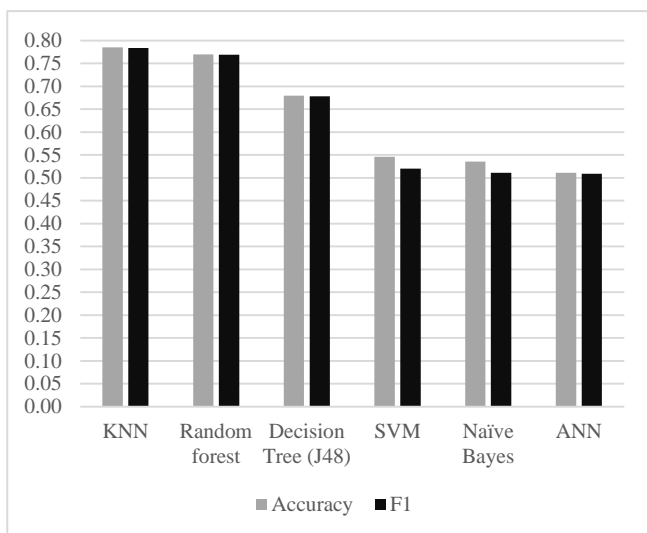


Fig. 2. Accuracy and F1 results.

In the preceding Table III and Fig. 2, we observe that Random Forest and KNN achieve the best results. KNN achieves the best results where accuracy = 78.474%, and F1 = 0.784. Next comes Random Forest with accuracy = 76.986%, and F1= 0.769. Decision Tree achieves good results with accuracy = 67.976%, and F1 = 0.678. SVM, Naïve Bayes, and ANN achieve low performance. SVM achieves accuracy = 54.58%, and F1 = 0.520. Then comes Naïve Bayes with low results of accuracy = 53.526%, and F1 = 0.511. Last comes ANN with the lowest results: accuracy = 51.093%, and F1 = 0.509. For time complexity, Random Forest, KNN, Naïve Bayes, and Decision Tree take seconds while SVM takes minutes. ANN takes more than two hours.

## VII. CONCLUSIONS AND FUTURE WORK

This study compared several popular machine learning algorithms for use in crime prediction, including KNN, Random Forest, SVM, ANN, Nave Bayes, and Decision Tree.

Our findings show that these algorithms can provide useful insights into predicting crime patterns, with KNN having the highest overall accuracy (78.474%) and F1 scores. The performance of each algorithm, however, varied depending on the dataset and crime type being analyzed.

According to our findings, using machine learning for crime prediction has the potential to improve public safety and law enforcement efforts. However, it is critical to recognize the limitations and ethical concerns associated with the use of predictive algorithms in criminal justice systems. Because machine learning models are only as good as the data on which they are trained, it is critical to ensure that crime prediction datasets are diverse and representative of the population. Furthermore, it is critical to address potential biases and avoid discrimination when deploying these models.

Also, by using the correlation, we discovered that adding more attributes, and detaching and elaborating the current data rather than grouping it into periods, may yield better results in the future.

Overall, our research highlights the potential and challenges of using machine learning to predict crime in Dubai. As the field develops, it will be critical to carefully evaluate and refine these algorithms to ensure their accuracy, fairness, and ethical implementation.

We suggest that future work in this area include more variables, such as: data about buildings, street names, exact locations containing longitude and latitude, data about the victims, income, and the relationship between the criminal and victim.

## REFERENCES

- [1] Population and Vital Statistics. Dubai Statistics Center. (n.d.). Retrieved January 22, 2023, from <https://www.dsc.gov.ae/en-us/Themes/Pages/Population-and-Vital-Statistics.aspx?Theme=42>
- [2] Police, D. (2023, January 13). Major Crime Statistics. Dubai Police . Retrieved January 15, 2023, from <https://www.dubaipolice.gov.ae/wps/portal/home/opendata/majorcrimestatistics>
- [3] WAM. (2022, October 15). Dubai Police records 65% drop in criminal reports during Q3. Wam. <https://www.wam.ae/en/details/1395303092140>

- [4] Bappee, F.K., Soares, A., Petry, L.M. et al. Examining the impact of cross-domain learning on crime prediction. *J Big Data* 8, 96 (2021). <https://doi.org/10.1186/s40537-021-00489-9>
- [5] Sattar, Abdus. (2021). Crime Rate Prediction Using Machine Learning and Data Mining. DOI:10.1007/978-981-15-7394-1\_5
- [6] Lenin Mookiah||William Eberle||Ambareen Siraj (2015). Survey of Crime Analysis and Prediction. Proceedings of the Twenty-Eighth International Florida Artificial Intelligence Research Society Conference (FLAIRS 2015).
- [7] Sathyadevan, Shiju & S., Devan & Gangadharan, Surya. (2014). Crime Analysis and Prediction Using Data Mining. 10.1109/CNSC.2014.6906719.
- [8] X. Zhang, L. Liu, L. Xiao and J. Ji, "Comparison of Machine Learning Algorithms for Predicting Crime Hotspots," in *IEEE Access*, vol. 8, pp. 181302-181310, 2020, doi: 10.1109/ACCESS.2020.3028420.
- [9] Stalidis, Panagiotis & Semertzidis, Theodoros & Daras, Petros. (2021). Examining Deep Learning Architectures for Crime Classification and Prediction. *Forecasting*, 3, 741-762. 10.3390/forecast3040046.-
- [10] Sattar, Abdus. (2021). Crime Rate Prediction Using Machine Learning and Data Mining. DOI:10.1007/978-981-15-7394-1\_5
- [11] O. Llahá, "Crime Analysis and Prediction using Machine Learning," 2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO), Opatija, Croatia, 2020, pp. 496-501, doi: 10.23919/MIPRO48935.2020.9245120.
- [12] Nagamani, Sikhinam & Bhavishya, & Kumar, Mr & Sree, T & Reddy, Lakireddy. (2022). An experimental study of Crime Prediction using Machine Learning Algorithms. *Test Engineering and Management*. 83. 17819 - 17825.
- [13] X. Zhang, L. Liu, L. Xiao and J. Ji, "Comparison of Machine Learning Algorithms for Predicting Crime Hotspots," in *IEEE Access*, vol. 8, pp. 181302-181310, 2020, doi: 10.1109/ACCESS.2020.3028420.
- [14] Alves, Luiz & Valentin Ribeiro, Haroldo & Rodrigues, Francisco. (2017). Crime prediction through urban metrics and statistical learning. <https://doi.org/10.1016/j.physa.2018.03.084>
- [15] McClendon, Lawrence & Meghanathan, Natarajan. (2015). Using Machine Learning Algorithms to Analyze Crime Data. *Machine Learning and Applications: An International Journal*. 2. 1-12. 10.5121/mlaij.2015.2101.
- [16] Major Crime Statistics. (n.d.). Dubai Police. Retrieved February 8, 2023, from <https://www.dubaipolice.gov.ae/wps/portal/home/opendata/majorcrimestatistics>
- [17] Narain, Profbhavana. (2021). An Empirical Analysis of Machine Learning Algorithms For Crime Prediction using Stacked Generalization: An Ensemble Approach. *IEEE Access*. XX. 1-9. 10.1109/ACCESS.2021.3075140,.
- [18] Azhari, Mourad & Alaoui, Altaf & Acharoui, Zakia & Ettaki, Badia & Zerouaoui, Jamal. (2019). Adaptation of the random forest method: solving the problem of pulsar search. *SCA '19: Proceedings of the 4th International Conference on Smart City Applications*. 1-6. 10.1145/3368756.3369004.
- [19] Cutler, Adele & Cutler, David & Stevens, John. (2011). *Random Forests*. 10.1007/978-1-4419-9326-7\_5.
- [20] Ali, Jehad & Khan, Rehanullah & Ahmad, Nasir & Maqsood, Imran. (2012). *Random Forests and Decision Trees*. *International Journal of Computer Science Issues(IJCSI)*. 9.
- [21] Oshiro, Thais & Perez, Pedro & Baranauskas, José. (2012). How Many Trees in a Random Forest?. *Lecture notes in computer science*. 7376. 10.1007/978-3-642-31537-4\_13.
- [22] Guo, Gongde & Wang, Hui & Bell, David & Bi, Yaxin. (2004). *KNN Model-Based Approach in Classification*.
- [23] Taunk, Kashvi & De, Sanjukta & Verma, Srishti & Swetapadma, Aleena. (2019). A Brief Review of Nearest Neighbor Algorithm for Learning and Classification.1255-1260.10.1109/ICCS45141.2019.9065747.
- [24] Md Imran Hossain. (2022). *Support Vector Machine\**. Frankfurt University of Applied Sciences. Frankfurt. Research for Master of Science in High Integrity Systems.
- [25] Berk, Richard. (2020). *Support Vector Machines*. 10.1007/978-3-030-40189-4\_7.
- [26] Bell, Jason. (2015). *Support Vector Machines*. 10.1002/9781119183464.ch7
- [27] Akgül, İsmail & Kaya, Volkan. (2022). A REVIEW ON ARTIFICIAL NEURAL NETWORKS.[https://www.researchgate.net/publication/360967369\\_A\\_REVIEW\\_ON\\_ARTIFICIAL\\_NEURAL\\_NETWORKS](https://www.researchgate.net/publication/360967369_A_REVIEW_ON_ARTIFICIAL_NEURAL_NETWORKS)
- [28] Yang, X.. (2009). Artificial neural networks. *Handbook of Research on Geoinformatics*. 122-128. 10.4018/978-1-59140-995-3.ch016.
- [29] Wang, Haonan & Chen, Yijia. (2022). Application of Artificial Neural Networks in Chemical Process Control. *Asian Journal of Research in Computer Science*. 22-37. 10.9734/ajrcos/2022/v14i130325.
- [30] Sai, Mitra & Kamasani, Sai Mitra. (2021). A STUDY ON NAIVE BAYES CLASSIFIER. [https://www.researchgate.net/publication/356267142\\_A\\_STUDY\\_ON\\_NAIVE\\_BAYES\\_CLASSIFIER](https://www.researchgate.net/publication/356267142_A_STUDY_ON_NAIVE_BAYES_CLASSIFIER)
- [31] Afdhaluzzikri, Afdhaluzzikri & Mawengkang, Herman & Sitompul, Opim. (2022). Performance analysis of Naive Bayes method with data weighting. *Sinkron*. 7. 817-821. 10.33395/sinkron.v7i3.11516.
- [32] Bshouty, Nader & Haddad-Zaknoon, Catherine. (2021). On Learning and Testing Decision Tree. <https://doi.org/10.48550/arXiv.2108.04587>
- [33] Jijo, Bahzad & Mohsin Abdulazeze, Adnan. (2021). Classification Based on Decision Tree Algorithm for Machine Learning. *Journal of Applied Science and Technology Trends*. 2. 20-28.
- [34] Nahler, Gerhard. (2010). Pearson correlation coefficient. 10.1007/978-3-211-89836-9\_1025.

# Preserving Cultural Heritage Through AI: Developing LeNet Architecture for Wayang Image Classification

Muhathir<sup>1</sup>, Nurul Khairina<sup>2</sup>, Rehia Karenina Isabella Barus<sup>3</sup>, Mutammimul Ula<sup>4</sup>, Ilham Sahputra<sup>5</sup>

Universitas Medan Area, Fakultas Teknik, Prodi Teknik Informatika, Medan, Indonesia<sup>1,2</sup>

Universitas Medan Area, Program Studi Ilmu Komunikasi, Fakultas Ilmu Sosial dan Ilmu Politik, Medan, Indonesia<sup>3</sup>

Universitas Malikussaleh, Sistem Informasi, Fakultas Teknik, Aceh, Indonesia<sup>4,5</sup>

**Abstract**—Wayang, an ancient cultural tradition in Java, has been an integral part of Indonesian culture for 1500 years. Rooted in Hindu cultural influences, wayang has evolved into a highly esteemed and beloved performance art. In the form of wayang kulit, this tradition conveys profound philosophical messages and implicit meanings that resonate with Javanese society. This research aims to develop an artificial intelligence (AI) model using deep learning with the LeNet architecture to accurately classify wayang images. The model was tested with 2515 Punakawan wayang images, showing excellent performance with an accuracy of 80% to 85%. Although the model successfully recognizes and distinguishes wayang classes, it faces some challenges in classifying specific classes, particularly in scenarios 2 and 4. Nevertheless, this research has a positive impact on cultural preservation, as the developed AI model can be used for automatic wayang image recognition. These implications open opportunities to better understand and preserve this rich cultural heritage through AI technology. With further improvements, this model has the potential to become a valuable tool in the efforts to preserve and introduce wayang culture to future generations.

**Keywords**—Wayang; LeNet; artificial intelligence; deep learning; cultural tradition

## I. INTRODUCTION

Wayang, a cultural tradition in Java, has been cherished and recognized by the Javanese people for around 1500 years. Its origins can be traced back to the influences of Hindu culture, particularly in the form of shadow puppetry, which eventually evolved into the renowned wayang performances [1], [2]. Indonesia is rich in folktales inherited from our ancestors, and among them, wayang kulit holds a special place as one of the most beloved and influential stories. Wayang kulit art is highly esteemed, primarily due to its profound philosophical values deeply rooted in the history of wayang kulit [3]. Wayang can be classified into two main types: wayang orang, performed live by human actors, and wayang boneka, controlled by a puppeteer called a dalang. Wayang kulit, a form of puppetry, features intricately crafted wooden puppets dressed in leather attire [4]–[8].

In Indonesia, the term "wayang" generally refers to puppetry [9]. The Javanese language defines "wayang" as shadow, while in Malay it denotes shadow, vagueness, or even transcendence [10], [11]. Wayang represents the embodiment of human qualities, encompassing virtues, greed, and more [12]–[15]. For over a thousand years, wayang has been cherished and embraced by the Javanese community, carrying

implicit meanings that resonate with their local languages. It showcases classical stories and narratives, often derived from the Ramayana and Mahabharata, which have been adapted to Javanese culture while maintaining their Hindu-Indian origins [13], [16]. Wayang performances hold significant value as they go beyond mere entertainment, serving as a form of cultural art that imparts life lessons, education, and guiding principles for living [17].

Wayang holds a versatile nature that allows its application in various contexts [18], [19]. A wayang performance carries numerous valuable lessons and life principles. Notably, it serves as an educational and moral medium, particularly benefiting the younger generation [17]. Wayang encompasses a wide range of types and forms, each possessing its distinct characteristics, influenced by specific regions. In Javanese cultural traditions, wayang exhibits diverse variations and styles, including wayang Gareng, Batara Wisnu, Yudishtira, Werkudara, Arjuna, and others, exemplifying the richness of its cultural significance.

The rapid advancement of technology in Indonesia has had a detrimental impact on the preservation of cultural heritage [20]. Wayang, one of Indonesia's invaluable cultural treasures, has historically played a vital role in shaping character through its insightful advice and captivating stories. However, the visibility of wayang performances has significantly declined, primarily due to a waning interest among the audience [21], [22]. Consequently, younger generations are becoming increasingly unfamiliar with the names and significance of wayang characters.

This diminishing awareness of wayang and other cultural aspects poses a significant challenge in contemporary society. To address this issue, it is essential to explore innovative approaches and leverage technological advancements. Computerization emerges as a potential solution to educate and engage the community effectively. Utilizing techniques such as LeNet classification, computer-based methods can be employed to enhance the understanding and appreciation of various wayang types and other cultural elements. By leveraging technology, we can revive interest, foster cultural appreciation, and ensure the preservation of Indonesia's rich cultural heritage for future generations.

The study introduces a CNN-based approach for recognizing artificial tire-side pressure printing characters on tire surfaces. The method employs image pre-processing with an enhanced SSR algorithm, character localization and

segmentation using template matching, and character recognition utilizing an improved LeNet-5 network structure. Experimental results showcase impressive recognition accuracy, with 95.9% accuracy on the training set, 99.5% accuracy on the validation set, and 95.6% accuracy on the testing set [23].

We propose a novel fault diagnosis model for rotating machinery that overcomes the limitation of directly feeding one-dimensional data into convolutional neural networks. Our model employs a rainbow recursive plot (RRP) to convert vibration signals into two-dimensional color images, allowing for the capture of important fault information. By utilizing a LeNet-5-based CNN, we extract features from the converted images, enabling accurate fault diagnosis recognition. Experimental results on public datasets show a significant improvement in recognition accuracy, reaching 97.86% [24].

Motivated by the strengths and uniqueness of the LeNet method in object classification, this research aims to create a new classification model applied to the case of wayang identification. The selection of wayang as the research case is based on its intricate shapes and the similarities among different types of wayang, posing a challenge in accurately detecting objects with high accuracy.

The explicit objectives of this research are as follows:

- Classify wayang types based on images using the LeNet architecture.
- Evaluate the performance of the classification approach based on the LeNet architecture with depthwise separable method.

## II. EXPERIMENT

### A. Research Architecture

The research architecture built in this study is presented in Fig. 1.

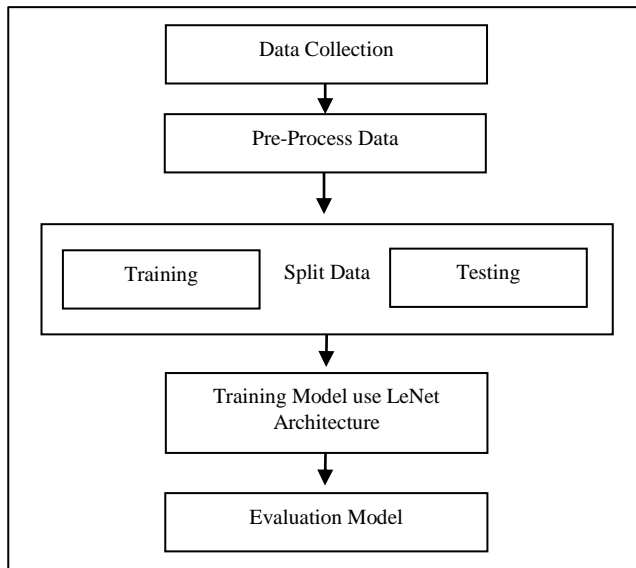


Fig. 1. Research architectural model utilized in this study for wayang classification.

Fig. 1 depicts the research architectural model utilized in this study for Wayang classification. The process begins with data collection, followed by pre-processing, where the dataset is curated to include suitable and unsuitable images as samples. Subsequently, the dataset is divided into three subsets: training, validation, and testing. The LeNet architecture is then employed to construct a deep learning model, utilizing the training set for model training. After the training phase, the model's performance and accuracy are evaluated using the validation set. The evaluation process involves the application of Eq. (1) to (4) to calculate relevant metrics such as accuracy, precision, recall, and F1-score. Finally, the model's effectiveness is assessed by testing it on the independent testing set. This comprehensive approach ensures the robustness and reliability of the developed Wayang classification model.

### B. Data Collection and Pre-Processing Data

In this study, the technique of collecting wayang data was conducted using a smartphone camera at a distance of approximately 30cm. The data collection procedure began by preparing the necessary equipment, which included a high-quality smartphone with a camera. The lighting conditions around the wayang object were then adjusted to ensure adequate lighting. Next, the distance between the smartphone camera and the wayang object was set at around 30cm to ensure optimal focus and capture detailed images. The position of the smartphone camera was also considered to align with the wayang object, avoiding perspective distortion. The camera settings on the smartphone were adjusted according to the needs, including image resolution and appropriate modes. During the photo capture process, the smartphone camera was directed directly at the wayang object without obstructing the light or object with hands or fingers. The captured photos were then examined to ensure clarity and detailed depiction of the wayang images. If necessary, the photo capture process could be repeated to obtain the best results. Thus, the technique of collecting wayang data using a smartphone camera at a distance of 30cm can provide adequate data for further analysis.

After collecting the wayang image data using a smartphone camera, the next step is to preprocess the data to prepare it for further analysis. Wayang data preprocessing involves several important stages. Firstly, the image format and resolution are adjusted to meet the analysis requirements. This includes resizing the images, adjusting the file format, and enhancing image clarity. In this preprocessing stage, efforts are made to remove noise or disturbances that may appear in the images. If necessary, disruptive elements such as shadows or other artifacts are also removed. By carefully following the preprocessing steps, the wayang image data is ready to be used in the next stage of analysis, such as pattern recognition and character classification of wayang. Thorough and meticulous preprocessing ensures the reliability and quality of the data that will be used in this research.

### C. Split Data

In the data analysis of this research, we used a sample of 2515 Punakawan wayang images, consisting of four different types of wayang. The sample was divided proportionally for

training and testing purposes. The training data, which accounts for 80% or approximately 2000 images, was used to train the model for analysis. The testing data, comprising 20% or about 515 images, was used to assess the model's performance and evaluate its ability to classify wayang images. This division is crucial to ensure a good representation of various types of wayang in the dataset and minimize bias in the analysis. Therefore, this data analysis provides a strong foundation for this research and ensures the validity of the results obtained from the developed model.

**D. LeNet Architecture**

LeNet-5 is a convolutional neural network (CNN) architecture designed for image recognition tasks. It was developed by Yann LeCun et al. in 1998 and has become a foundational model in deep learning [25]–[27]. The architecture consists of convolutional layers, pooling layers, and fully connected layers. For LeNet-5, the input is typically a grayscale or color image with a size of 32x32 pixels. The convolutional layers apply learnable filters to extract features from the input image, generating feature maps that capture different aspects of the image. Pooling layers reduce the spatial dimensions of the feature maps, aiding in feature extraction and computational efficiency. Max pooling is commonly used, selecting the maximum value within a specified window. The resulting feature maps are then flattened and fed into fully connected layers for classification. The final fully connected layer produces the output, representing the predicted class probabilities. LeNet-5 has played a significant role in advancing CNNs and remains a widely studied architecture in image recognition. Fig. 2 illustrate the architectural design of LeNet, which was developed for the purpose of classifying wayang.

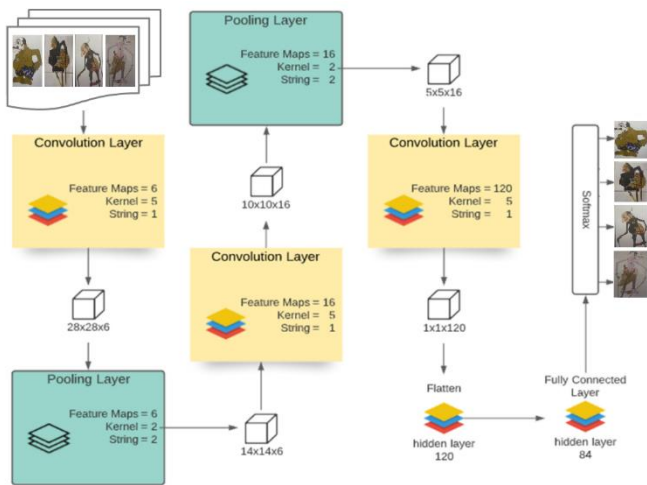


Fig. 2. LeNet architecture.

**E. Hyperparameter Initialization**

Learning parameters play a crucial role in the accuracy and performance of a model during training. This study explores several important learning parameters, namely Epoch, Batch size, Optimizer, and activation functions. Epoch determines the number of times the entire dataset is used during training. A higher epoch value increases the number of iterations but may lead to overfitting. Batch size refers to the number of data

samples used in each iteration. A larger batch size enhances training speed but may impact generalization. Optimizer algorithms like SGD, Adam, and RMSprop optimize the model during training, each with its strengths and weaknesses. Activation functions introduce non-linearity, such as ReLU, Sigmoid, and Tanh, influencing convergence speed and feature representation. Selecting appropriate learning parameters is crucial for achieving optimal model performance. Careful experimentation and parameter tuning are necessary to find the most suitable combination for each specific task and dataset. The candidate hyperparameters utilized in this study are outlined in Table I.

TABLE I. HYPERPARAMETER MODEL

Parameter	Candidate
Epoch	(10, 50, and 100)
Batch Size	(10,20,40,60,80,and 100)
Optimizer	(SGD, RMSprop, Adagrad, Adam and Nadam )
Activation function	(Tanh, Relu, Linear, Sigmoid, Softmax)

**F. Performa Measure**

The confusion matrix is a valuable method for evaluating the accuracy of an object estimation model. It compares the predicted classification results with the actual classes and provides a detailed view of the model's performance [28]–[30]. The accuracy of the method reflects how well the predicted values align with the actual values. Precision, on the other hand, measures the repeatability of the measurements or the proportion of accurate predictions. Recall represents the number of correct positive responses identified by the model. Combining precision and recall yields the f1-score, which gives a balanced average assessment of the model's performance. These metrics can be calculated using the following formulas, where TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative, respectively [31]–[33].

Description

TP = True Positive

FP = False Positive

FN = False Negative

TN = True Negative

$$Accuracy = \frac{TN+TP}{TN+FP+TP+FN} \tag{1}$$

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

$$F1 = \frac{2*Presicion*Recall}{Presicion+Recall} \tag{4}$$

**III. RESULT AND DISCUSSION**

In this session, we present the research results on wayang classification using the deep learning architecture LeNet. Two scenarios were built to extract information from object images, namely using the default LeNet and implementing the

depthwise separable method. Depthwise separable is a data processing technique in neural networks that combines two convolution stages: depthwise convolution and pointwise convolution. Depthwise convolution applies filters to each input channel separately to reduce the number of parameters [34], [35]. Furthermore, we conducted a search for the best hyperparameters to improve model performance using Grid search. This approach aims to achieve optimal model performance by selecting the right combination of hyperparameters, enabling the model to provide accurate and efficient results in the wayang classification training and testing process.

### A. Samplel Wayang

Fig. 3 illustrate the unique characteristics of each wayang character. In this image, Bagong is seen with a cheerful face full of wit, while Gareng appears loyal and friendly with a heartwarming smile. Additionally, Petruk's image portrays a funny yet clumsy figure, bringing joy with his comical expressions. On the other hand, the image of Semar presents a wise and gentle character with a gaze full of understanding. Through these images, the distinctiveness and roles of each character in wayang performances can be clearly explained, enhancing the charm and allure of Indonesia's wayang art. The beauty and profound meaning depicted in Fig. 3 enables the audience to appreciate and understand the local wisdom and cultural richness of Indonesia that emanate from this traditional wayang art.

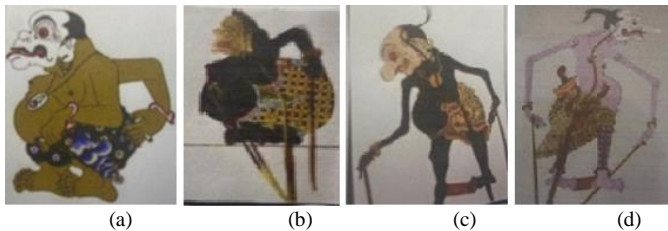


Fig. 3. Indonesian wayang characters (a) Bagong, (b) Gareng, (c) Petruk, and (d) Semar.

### B. Training LeNet

The deep learning architecture based on LeNet is designed for image classification, as depicted in Fig. 2. The input is a 32x32x1 image, with the last dimension representing the color channel. The network includes two convolutional layers with 28 and 10 filters, followed by 2x2 max pooling layers to reduce feature map size. Rectified Linear Unit (ReLU) activation functions are used in the convolutional layers. After the second pooling layer, the output feature maps are flattened into a 120-dimensional vector. Dropout is applied to prevent overfitting by randomly zeroing input units during training. The dropout output then goes through a fully connected layer with 84 neurons using the dense activation function. Batch normalization is employed to normalize activations and improve model stability. ReLU is a popular activation function for its efficiency and performance improvement. Max pooling down samples feature maps to extract essential characteristics and enhance the network's effectiveness. A flatten layer transforms the output of the last convolutional layer into a 1D vector, followed by a fully connected dense layer. The final classification is determined by this layer, using the softmax

activation function to generate probability scores for each class. The output layer produces the final classification predictions.

Based on the results of training the LeNet model on the wayang classification task, as presented in Fig. 4, a total of 61,496 parameters were obtained. These parameters encompass all the weights and biases that are fine-tuned and optimized throughout the training process. The overall count of parameters serves as an indicator of the model's complexity and capacity to discern patterns and features from the data. A greater number of parameters empowers the model to acquire intricate data representations. However, this must be carefully balanced against the potential for overfitting and the increased computational demands.

```
Model: "sequential_7"
-----
```

Layer (type)	Output Shape	Param #
conv2d_14 (Conv2D)	(None, 28, 28, 6)	456
max_pooling2d_14 (MaxPooling2D)	(None, 14, 14, 6)	0
conv2d_15 (Conv2D)	(None, 10, 10, 16)	2416
max_pooling2d_15 (MaxPooling2D)	(None, 5, 5, 16)	0
flatten_7 (Flatten)	(None, 400)	0
dense_21 (Dense)	(None, 120)	48120
dense_22 (Dense)	(None, 84)	10164
dense_23 (Dense)	(None, 4)	340

```
-----
Total params: 61,496
Trainable params: 61,496
Non-trainable params: 0
```

Fig. 4. Training model LeNet (scenario 1).

The evaluation of deep learning models during training is primarily based on two key metrics: training loss and validation loss. Training loss measures the discrepancy between the predicted and actual values on the training data, while validation loss measures the discrepancy on the validation data. It is crucial to analyze both training and validation loss to obtain a comprehensive understanding of the model's performance. For visualizing the performance of the proposed algorithm, Fig. 5 presents the performance visualization, which incorporates both training and validation loss.

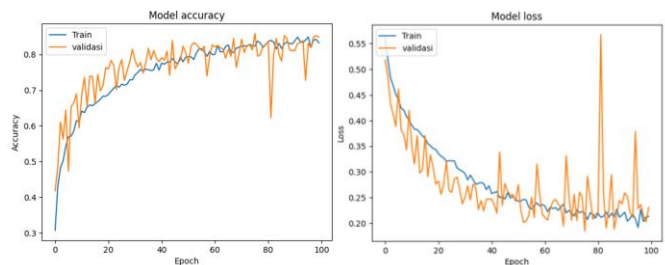


Fig. 5. Training and validation model LeNet (Scenario 1).

C. Training LeNet with Depthwise Separable

In this scenario, the deep learning model was constructed with the same architecture as the previous model, but with the integration of the depthwise separable approach to effectively diminish the total number of parameters. Following the application of the depthwise separable approach, the entire model comprises a total of 58,985 parameters. This represents a notable reduction in the parameter count compared to the previous model, all while maintaining the quality of the classification performance, as illustrated in Fig. 6.

Layer (type)	Output Shape	Param #
separable_conv2d (Separable Conv2D)	(None, 28, 28, 6)	99
max_pooling2d_2 (MaxPooling2D)	(None, 14, 14, 6)	0
separable_conv2d_1 (Separable Conv2D)	(None, 10, 10, 16)	262
max_pooling2d_3 (MaxPooling2D)	(None, 5, 5, 16)	0
flatten_1 (Flatten)	(None, 400)	0
dense_3 (Dense)	(None, 120)	48120
dense_4 (Dense)	(None, 84)	10164
dense_5 (Dense)	(None, 4)	340

=====  
 Total params: 58,985  
 Trainable params: 58,985  
 Non-trainable params: 0

Fig. 6. Training model LeNet with depthwise separable (scenario 2).

For visualizing the performance of the proposed algorithm, Fig. 7 presents the performance visualization, which incorporates both training and validation loss.

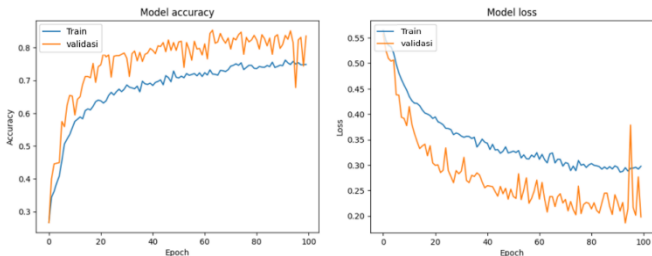


Fig. 7. Training and validation model LeNet with depthwise separable (Scenario 2).

D. Hyperparameter tuning

In this scenario, we used grid search to tune the hyperparameters of the LeNet model for both Scenario 1 and Scenario 2. Grid search is a systematic technique that allows us to explore various combinations of hyperparameter values. By experimenting with different combinations, we can find the best hyperparameter settings in terms of model performance and generalization for the wayang classification task. The results of tuning the hyperparameters for each scenario are presented in Table II. Additionally, to visualize the training and validation performance of the model, the training and

validation curves for each scenario are shown in Fig. 8 and Fig. 9, respectively. These graphs provide a visual representation of how the model learns and adapts during the training process.

TABLE II. GRID SEARCH HYPERPARAMETER TUNING RESULTS

Parameter	Candidate	Select Hyperparameter in Scenario 1	Select Hyperparameter in Scenario 2
Epoch	(10, 50, dan 100)	100	100
Batch Size	(10,20,40,60,80,dan 100)	10	10
Optimizer	(SGD, RMSprop,Adagrad, Adam dan Nadam )	Adam	SGD
Fungsi Aktivasi	(Tanh, Relu, Linear, Sigmoid, Softmax)	Relu	Relu

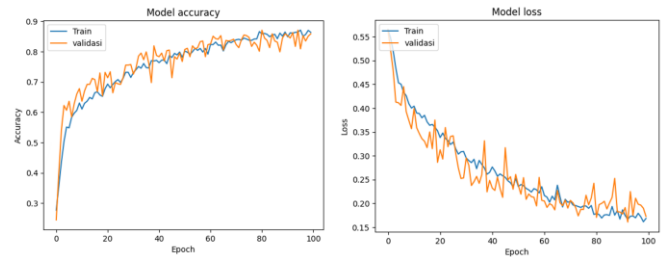


Fig. 8. Training model hyperparameter Scenario 1 (Scenario 3).

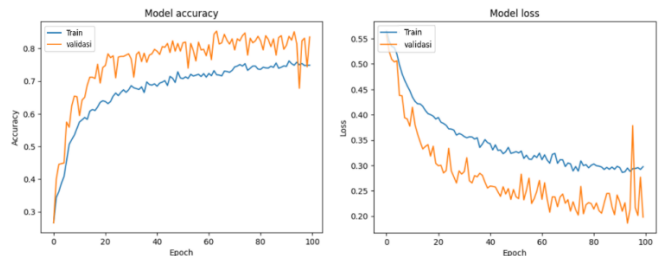


Fig. 9. Training model hyperparameter Scenario 2 (Scenario 4).

E. Evaluation

The findings of four scenarios utilizing the confusion matrix are depicted in Fig. 10, providing a valuable tool for evaluating model performance in classification tasks. The study employed the confusion matrix to assess four distinct scenarios' outcomes in data classification. By utilizing this matrix, we gauge the model's accuracy in classifying data accurately. Information from the confusion matrix aids in calculating essential performance evaluation metrics like precision, recall, accuracy, and F1-Score.

Furthermore, the performance metrics for the four scenarios are presented in Tables III-VI, offering a comprehensive evaluation of the model's classification performance. These tables present precision, recall, accuracy, and F1-Score for each scenario, crucial for assessing the model's effectiveness. Table III corresponds to Scenario 1, Table IV to Scenario 2, Table V to Scenario 3, and Table VI to Scenario 4. Analyzing these metrics allows us to determine the model's performance in various scenarios, facilitating comparisons to identify the most suitable approach for the classification task at hand. These tables play a pivotal role in comprehending the strengths



and weaknesses of each scenario, guiding data-driven decisions to optimize and enhance the model's classification performance. The confusion matrix analysis and performance metrics' results provide valuable insights into the model's capabilities, enabling us to refine and fine-tune it for improved classification results.

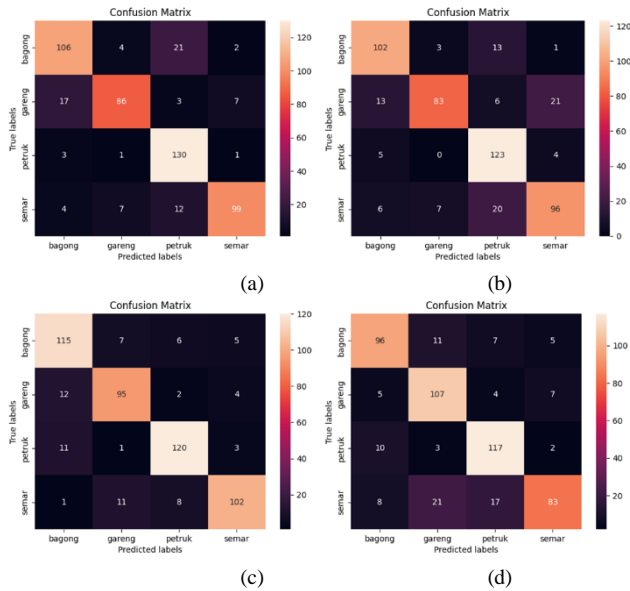


Fig. 10. Confusion Matrix Results (a) Scenario 1, (b) Scenario 2, (c) Scenario 3, and (d) Scenario 4.

TABLE III. LEnET PERFORMANCE IN SCENARIO 1

	Precision	Recall	F1_score
<b>Bagong</b>	0.81	0.79	0.80
<b>Gareng</b>	0.87	0.76	0.81
<b>Petruk</b>	0.78	0.96	0.86
<b>Semar</b>	0.90	0.81	0.85
<b>Accuracy</b>	0.83		

TABLE IV. LEnET PERFORMANCE IN SCENARIO 2

	Precision	Recall	F1_score
<b>Bagong</b>	0.80	0.85	0.83
<b>Gareng</b>	0.89	0.67	0.76
<b>Petruk</b>	0.75	0.93	0.83
<b>Semar</b>	0.78	0.74	0.76
<b>Accuracy</b>	0.80		

TABLE V. LEnET PERFORMANCE IN SCENARIO 3

	Precision	Recall	F1_score
<b>Bagong</b>	0.83	0.87	0.85
<b>Gareng</b>	0.90	0.76	0.82
<b>Petruk</b>	0.90	0.93	0.92
<b>Semar</b>	0.81	0.87	0.84
<b>Accuracy</b>	0.85		

TABLE VI. LEnET PERFORMANCE IN SCENARIO 4

	Precision	Recall	F1_score
<b>Bagong</b>	0.73	0.84	0.78
<b>Gareng</b>	0.77	0.74	0.76
<b>Petruk</b>	0.88	0.84	0.86
<b>Semar</b>	0.82	0.78	0.80
<b>Accuracy</b>	0.80		

In Scenario 1, the model achieved an accuracy of 83%. The evaluation results showed good performance in classifying all wayang classes. Petruk had the highest precision and recall values, indicating its excellent ability to identify the Petruk class. Bagong and Semar also had good F1\_score values, indicating a balance between precision and recall. However, it should be noted that the precision for Petruk and Semar classes was slightly lower than recall, possibly due to some difficulty in classifying complex samples.

In Scenario 2, the model achieved an accuracy of 80%. There was variation in performance among the wayang classes. Gareng had a lower recall value, indicating some difficulty in accurately recognizing the Gareng class. Bagong and Petruk had good F1\_score values, but the precision for the Petruk class was lower than recall. Semar had low F1-score values, suggesting challenges in classifying the Semar class accurately.

In Scenario 3, the model achieved an accuracy of 85%. The model showed consistent performance for all wayang classes with high F1\_score values. The Petruk class had the highest precision and recall values, indicating an excellent ability to classify this class. The Gareng class also had a good F1\_score. These results indicate that the model has a strong ability to recognize and distinguish between different wayang classes.

In Scenario 4, the model achieved an accuracy of 80%. Precision and recall for all wayang classes were relatively balanced, but the F1\_score for the Gareng and Semar classes was slightly lower than other classes. It is important to note that the model showed some difficulty in accurately classifying the Gareng and Semar classes.

Overall, the evaluation results indicate that the model has good performance in classifying the wayang classes. However, there is some variation in performance among specific classes, which could be the focus of further model improvements. Careful evaluation of all performance metrics can help identify areas where the model can be enhanced to achieve better results in wayang classification tasks.

F. Discussion

The evaluation results of the four wayang classification scenarios show good performance in classifying wayang images. All scenarios achieve relatively high accuracy, ranging from 80% to 85%. Additionally, the precision, recall, and F1-Score values are balanced, indicating the model's ability to classify wayang images accurately and consistently. The best scenario is Scenario 3, which achieves an accuracy of 85% with a high precision of 86%, recall, and F1-Score of 85.75%. This demonstrates that the model in this scenario excels in classifying wayang images with great accuracy. On the other hand, the worst scenario is Scenario 2, with an accuracy of 80% and slightly lower precision, recall, and F1-Score compared to the other scenarios. Although it still performs well, this scenario can be improved to achieve higher accuracy and consistency. Overall, the evaluation results indicate that the developed model performs well in the wayang classification task and is reliable for further analysis. The best and worst scenarios can serve as a basis for further improvements and the development of more optimal models in wayang image recognition.

The AI models developed using the deep learning approach with LeNet (Scenarios 1, 2, 3, and 4) exhibited better performance compared to AI models using other methods such as KNN + GLCM [5] and MLP + GLCM [36]. The LeNet-based AI models achieved higher accuracy and more balanced precision, recall, and F1-Score, as shown in Table VII.

The results of this study provide insight into the effectiveness of using depthwise separable LeNet models in image-based puppet classification. This contributes to the development of better classification techniques in the domain of puppet recognition. Moreover, the use of depthwise separable is able to reduce the total number of model parameters, which indicates efficiency in the use of computational resources, although the addition of depthwise separable affects the level of accuracy gain.

In future research, there are several recommendations that can be considered. First, increasing the amount of puppet image data can help improve the performance of the model. Second, further exploration of hyperparameters and optimization techniques can be an important step in improving classification accuracy such as random search and Bayesian optimization. In addition, future research can consider the use of attention modules to improve the model's ability to deal with a larger variety of wayang images.

TABLE VII. COMPARISON WITH PREVIOUS STUDIES

Study (Ref)	AI Model	Accuracy	Precision	Recall	F1-Score
Sandy et al [5]	KNN + GLCM	0.775	-	-	-
Muhathir et al [30]	SVM + GLCM	0.834	-	-	-
Santoso et al [36]	MLP + GLCM	0.734	-	-	-
Scenario 1	Lenet	0.83	0.84	0.83	0.83
Scenario 2	Lenet	0.8	0.805	0.7975	0.795
Scenario 3	Lenet	0.85	0.86	0.8575	0.8575
Scenario 4	Lenet	0.8	0.8	0.8	0.8

#### IV. CONCLUSION

This research contributes to the development of AI models for wayang image classification using deep learning with the LeNet architecture. The AI models developed show superior performance in classifying wayang images compared to models using other methods such as KNN + GLCM and MLP + GLCM.

The evaluation results indicate that all wayang classification scenarios achieve relatively high accuracy, ranging from 80% to 85%. The precision, recall, and F1-Score values are well-balanced, demonstrating the model's ability to classify wayang images accurately and consistently. However, there is variation in performance among specific wayang classes, suggesting potential areas for further model improvements. Careful evaluation of all performance metrics

can help identify these areas and enhance the model's performance in wayang classification tasks.

The research has the potential to benefit cultural recognition and preservation, as the developed AI model can be used for further analysis and automated recognition of wayang images. However, the study has some limitations, such as the size of the dataset and the complexity of wayang images, which may affect model performance. Therefore, future research should consider augmenting the data and exploring alternative deep learning approaches or combining different methods to further enhance the model's accuracy and robustness in wayang image recognition.

#### REFERENCES

- [1] B. Anggoro, "Wayang dan Seni Pertunjukan: Kajian Sejarah Perkembangan Seni Wayang di Tanah Jawa sebagai Seni Pertunjukan dan Dakwah," *JUSPI: Jurnal Sejarah Peradaban Islam*, vol. 2, no. 2, 2018.
- [2] A. Purwantoro, N. S. Prameswari, and R. B. M. N. Mohd Nasir, "The Development of the Indonesian Culture Gunung Design: Wayang Godhong 'Smoking Violated,'" *Harmonia: Journal of Arts Research and Education*, vol. 22, no. 1, pp. 62–77, Jun. 2022, doi: 10.15294/harmonia.v22i1.36525.
- [3] E. Nurcahyawati and M. Arifin, "Manifestasi Transformasi Nilai-Nilai Ajaran Islam Dalam Tokoh Wayang Kulit Pandawa Lima Pada Cerita Mahabharata," *Jurnal Dirosah Islamiyah*, vol. 4, p. 304, 2022, doi: 10.17467/jdi.v4i2.1078.
- [4] M. Resa Arif Yudianto, K. Kusriani, and H. Al Fatta, "ANALISIS PENGARUH TINGKAT AKURASI KLASIFIKASI CITRA WAYANG DENGAN ALGORITMA CONVOLUTIONAL NEURAL NETWORK," *Jurnal Teknologi Informasi*, vol. 4, no. 2, 2020.
- [5] B. Sandy, J. K. Siahaan, P. Permana, and \* Muhathir, *Klasifikasi Citra Wayang Dengan Menggunakan Metode k-NN & GLCM*, vol. 2. 2019.
- [6] A. Susanto, I. Utomo, and W. Mulyono, "REKOGNISI WAYANG KULIT MENGGUNAKAN JARINGAN SYARAF TIRUAN," in *Prosiding SENDI\_U*, 2019.
- [7] A. Setya, S. Pratama, A. Prasetya Wibawa, and A. N. Handayani, "CONVOLUTIONAL NEURAL NETWORK (CNN) UNTUK MENENTUKAN GAGRAK WAYANG KULIT," 2022.
- [8] M. Muhathir, M. H. Santoso, and D. A. Larasati, "Wayang Image Classification Using SVM Method and GLCM Feature Extraction," *JOURNAL OF INFORMATICS AND TELECOMMUNICATION ENGINEERING*, vol. 4, no. 2, pp. 373–382, Jan. 2021, doi: 10.31289/jite.v4i2.4524.
- [9] M. I. Cohen, "The Reverse Repatriation of Javanese Puppets," *Theatre Journal*, vol. 69, no. 3, pp. 361–381, 2017, [Online]. Available: <https://www.jstor.org/stable/48560802>
- [10] D. Endah Ciswiyati, M. Ibban Syarif, and S. Muharrar, "Catharsis: Journal of Arts Education Creativity Overview: A Contemporary Wayang By Nanang Garuda," *Catharsis: Journal of Arts Education*, vol. 10, no. 2, pp. 171–180, 2021, doi: 10.15294/catharsis.v10i2.52632.
- [11] E. Nurcahyawati and M. Arifin, "Manifestasi Transformasi Nilai-Nilai Ajaran Islam Dalam Tokoh Wayang Kulit Pandawa Lima Pada Cerita Mahabharata," *Jurnal Dirosah Islamiyah*, vol. 4, p. 304, 2022, doi: 10.17467/jdi.v4i2.1078.
- [12] F. Reffiane, I. Mazidati, P. PGSD Universitas PGRI Semarang, and J. Sidodadi Timur No, "IMPLEMENTASI PENGEMBANGAN MEDIA WAYANG KERTON PADA TEMA KEGIATAN SEHARI-HARI," vol. 3, no. 2, pp. 163–170, 2016, doi: 10.17509/mimbar-sd.v3i2.4256.
- [13] E. Setiawan, "MAKNA FILOSOFI WAYANG PURWA DALAM LAKON DEWA RUCL," *Kontemplasi*, vol. 5, no. 2, 2017.
- [14] A. D. P. Putera, A. N. Hidayah, and A. Subiantoro, "Thermo-economic analysis of a geothermal binary power plant in Indonesia—a pre-feasibility case study of the Wayang Windu site," *Energies (Basel)*, vol. 12, no. 22, Nov. 2019, doi: 10.3390/en12224269.

- [15] M. Hynson, "A balinese 'call to prayer': Sounding religious nationalism and local identity in the Puja Tri Sandhya," *Religions (Basel)*, vol. 12, no. 8, Aug. 2021, doi: 10.3390/rel12080668.
- [16] T. Santoso and B. Setyawan, "Wayang Golek Menak: Wayang Puppet Show as Visualization Media of Javanese Literature," European Alliance for Innovation n.o., Oct. 2019. doi: 10.4108/eai.27-4-2019.2286930.
- [17] S. Purwanto, "Pendidikan Nilai dalam Pagelaran Wayang Kulit," *Ta'allum: Jurnal Pendidikan Islam*, vol. 6, no. 1, Mar. 2018, doi: 10.21274/taalum.2018.6.1.1-30.
- [18] M. Sakashita, T. Minagawa, A. Koike, I. Suzuki, K. Kawahara, and Y. Ochiai, "You as a Puppet: Evaluation of Telepresence User Interface for Puppetry," in *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, in UIST '17. New York, NY, USA: Association for Computing Machinery, 2017, pp. 217–228. doi: 10.1145/3126594.3126608.
- [19] Q. Fu and Q. Hu, "Study of Chinese Shadow Mapping Classification with the Application of Deep Learning Algorithms," *Comput Intell Neurosci*, vol. 2022, 2022, doi: 10.1155/2022/7050260.
- [20] B. P. Bangsa and L. H. Sihombing, "Impact of Japanese Popular Culture to Indonesian younger Generation:," *Humaniora*, vol. 13, no. 3, pp. 241–246, Nov. 2022, doi: 10.21512/humaniora.v13i3.8131.
- [21] J. Kurscheid *et al.*, "Shadow puppets and neglected diseases: Evaluating a health promotion performance in rural Indonesia," *Int J Environ Res Public Health*, vol. 15, no. 9, Sep. 2018, doi: 10.3390/ijerph15092050.
- [22] C. Williams *et al.*, "Shadow puppets and neglected diseases (2): A qualitative evaluation of a health promotion performance in rural Indonesia," *Int J Environ Res Public Health*, vol. 15, no. 12, Dec. 2018, doi: 10.3390/ijerph15122829.
- [23] Z. Guo, J. Yang, X. Qu, and Y. Li, "Fast Localization and High Accuracy Recognition of Tire Surface Embossed Characters Based on CNN," *Applied Sciences (Switzerland)*, vol. 13, no. 11, Jun. 2023, doi: 10.3390/app13116560.
- [24] X. Wang, X. Wang, T. Li, and X. Zhao, "A Fault Diagnosis Method Based on a Rainbow Recursive Plot and Deep Convolutional Neural Networks," *Energies (Basel)*, vol. 16, no. 11, Jun. 2023, doi: 10.3390/en16114357.
- [25] C. Chen, L. Jing, H. Li, Y. Tang, and F. Chen, "Individual Tree Species Identification Based on a Combination of Deep Learning and Traditional Features," *Remote Sens (Basel)*, vol. 15, no. 9, 2023, doi: 10.3390/rs15092301.
- [26] Z. Guo, J. Yang, X. Qu, and Y. Li, "Fast Localization and High Accuracy Recognition of Tire Surface Embossed Characters Based on CNN," *Applied Sciences*, vol. 13, no. 11, 2023, doi: 10.3390/app13116560.
- [27] N. Mao, H. Yang, and Z. Huang, "A Parameterized Parallel Design Approach to Efficient Mapping of CNNs onto FPGA," *Electronics (Basel)*, vol. 12, no. 5, 2023, doi: 10.3390/electronics12051106.
- [28] M. Melisah and M. Muhathir, "A modification of the Distance Formula on the K-Nearest Neighbor Method is Examined in Order to Categorize Spices from Photo Using the Histogram of Oriented Gradient \*," in *2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE)*, 2023, pp. 23–28. doi: 10.1109/ICCoSITE57641.2023.10127780.
- [29] I. Safira and M. Muhathir, "Analysis of Different Naïve Bayes Methods for Categorizing Spices Through Photo using the Speeded-up Robust Feature," in *2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE)*, 2023, pp. 29–34. doi: 10.1109/ICCoSITE57641.2023.10127787.
- [30] Muhathir and Al-Khowarizmi, "Measuring the Accuracy of SVM with Varying Kernel Function for Classification of Indonesian Wayang on Images," in *2020 International Conference on Decision Aid Sciences and Application, DASA 2020*, Institute of Electrical and Electronics Engineers Inc., Nov. 2020, pp. 1190–1196. doi: 10.1109/DASA51403.2020.9317197.
- [31] M. Muhathir, M. F. D. Ryandra, R. B. Y. Syah, N. Khairina, and R. Muliono, "Convolutional Neural Network (CNN) of Resnet-50 with Inceptionv3 Architecture in Classification on X-Ray Image," in *Artificial Intelligence Application in Networks and Systems*, P. Silhavy Radek and Silhavy, Ed., Cham: Springer International Publishing, 2023, pp. 208–221.
- [32] M. Ula, M. Muhathir, and I. Sahputra, "Optimization of Multilayer Perceptron Hyperparameter in Classifying Pneumonia Disease Through X-Ray Images with Speeded-Up Robust Features Extraction Method," *IJACSA) International Journal of Advanced Computer Science and Applications*, vol. 13, no. 10, 2022, [Online]. Available: www.ijacsa.thesai.org
- [33] Muhathir, R. A. Rizal, J. S. Sihotang, and R. Gultom, "Comparison of SURF and HOG extraction in classifying the blood image of malaria parasites using SVM," in *2019 International Conference of Computer Science and Information Technology (ICoSNIKOM)*, 2019, pp. 1–6. doi: 10.1109/ICoSNIKOM48755.2019.9111647.
- [34] Y. Guo, Y. Li, L. Wang, and T. Rosing, "Depthwise Convolution Is All You Need for Learning Multiple Visual Domains," in *The Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19)*, 2019. [Online]. Available: www.aaai.org
- [35] X. Liu, G. Yan, and Y. Chen, "Depthwise and spatial factorized network: A light-weight network for real-time semantic segmentation," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Sep. 2020. doi: 10.1088/1742-6596/1627/1/012023.
- [36] M. H. Santoso, D. A. Larasati, and M. Muhathir, "Wayang Image Classification Using MLP Method and GLCM Feature Extraction," *Journal of Computer Science, Information Technology and Telecommunication Engineering*, Sep. 2020, doi: 10.30596/jcositte.v1i2.5131.

# A Comprehensive Review of Modern Methods to Improve Diabetes Self-Care Management Systems

Alhuseen Omar Alsayed<sup>1</sup>, Nor Azman Ismail<sup>2</sup>, Layla Hasan<sup>3</sup>, Farhat Embarak<sup>4</sup>

School of Computing, Faculty of Engineering, University Teknologi Malaysia (UTM), Johor Bahru 81310, Malaysia<sup>1,2,3,4</sup>  
Department of Research Affairs Unit, Deanship of Scientific Research (DSR),  
King Abdulaziz University, Jeddah 21589, Saudi Arabia<sup>1</sup>

**Abstract**—Diabetes mellitus has become a global epidemic, with an increasing number of individuals affected by this chronic metabolic disorder. Effective management of diabetes requires a comprehensive self-care approach, which encompasses various aspects like monitoring blood glucose levels, adherence to medication, modifications in lifestyle, and regular healthcare monitoring. Innovative techniques for bettering diabetic self-care management have been developed recently as a result of developments in technology and healthcare systems. This comprehensive review examines the modern methods that have emerged to enhance diabetes self-care management systems. The review focuses on the integration of technology, Behavioural Change Techniques (BCTs), behavioural health theories such as Transtheoretical Model (TTM), the Health Belief Model (HBM), Theory of Reasoned Action/Planned Behaviour (TPB), Social Cognitive Theory (SCT) techniques to promote optimal diabetes care outcomes. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 standards were followed in this research's documentation. The Systematic Literature Review (SLR) period, which covered 2009 to 2020, was used to acquire the most recent complete review. Overall, the SLR results show that self-care interventions have a favourable impact on behaviours modification, the encouragement of good lifestyle habits, the lowering of blood glucose scales, and the accomplishment of significant weight loss. According to the review's findings, treatments for diabetic self-management that included behavioural health theories and BCTs in their creation tended to be more successful. In order to assist academics and practitioners with the creation of future applications, the restriction and future direction were finally defined. After recognising the potential for combining BCT methodologies and theories, it creates self-management interventions. Depending on these recognised cutting-edge mechanisms, the current SLR can assist application developers create a model to construct efficient self-care interventions for diabetes.

**Keywords**—Diabetes self-care; diabetes management; systematic literature review; BCT theories

## I. INTRODUCTION

Diabetes is a serious condition caused due to high blood sugar levels. High blood pressure, renal disease, and heart failure may all occur as a result of this chronic illness [1]–[4]. Type 1 diabetes and Type 2 diabetes represent two different types of diabetes. While Type 2 diabetes, which accounts for roughly 90% of all cases, is the most common type, Type 1 diabetes, which can't be prevented, is characterised by inadequate insulin in youngsters. A major issue faced by the health system is the rising incidence of diabetes, which has

turned into a global crisis. For example, according to the National Diabetes Statistics Report, 13% of Americans have diabetes. Global estimates indicate that 9% of people worldwide have diabetes, and that number is expected to rise to 12% by 2030 [5]. Ageing, obesity, and people's dietary habits are thought to be contributing factors to the rising number of diabetes [3]. These troubling trends highlight the critical requirement for researchers and technologists to develop practical diabetes treatment strategies. Several applications have been created for managing diabetes self-care as a result of the rise of diabetes cases in the world. Applications for managing diabetes are believed to be among the most popular ones from the Google Play store [4]. Numerous studies have demonstrated that diabetes self-care practises can considerably enhance a range of clinical outcomes for diabetics [5], [6]. In earlier studies, the effectiveness of diabetes self-care tools was qualitatively investigated [7], [8]. Additionally, a number of meta-analyses [9]–[11] have examined the usefulness of self-care tools for managing diabetes using quantitative data. The majority of previous investigations, nevertheless, ought to have uncovered the impact of BCTs on self-care management [11], [12]. Besides self-monitoring, the major problem with the existing system models is that the majority of the current mHealth apps make limited usage of BCTs and have few features [13], [14]. Moreover, the absence of customized feedback, poor user interface design, and accessibility concerns with current medical monitoring programmes are further problems (e.g. limited data entry options) [7], [15]–[17]. Additionally, diabetes patients think that e-self-management health applications should be interesting and provide a variety of functions that cover an extensive amount of information encompassing emotional and psychological support [18]. The impact of individual BCTs on diabetes management cannot be discounted because managing diabetes is strongly related to behaviour that requires appropriate modification [19]–[22]. Therefore, it seems like there is a lot of uncertainty if these mobile and web-based healthcare applications are an affordable means to give diabetes self-management education and whether they improve health outcomes and provide support in the real world [19], [20], [23].

The study aims to comprehensively review and analyze modern methods designed to enhance diabetes self-care management systems. The formulation of research questions (RQs) is crucial for determining the overarching goals and anticipated results of a study. The study mainly focuses on the research question “How can we use technology and

behaviour-based techniques to make diabetes self-care better and healthier for people with diabetes?" To find a solution for this question, following research questions are framed.

- 1) How well do diabetes management applications support and facilitate diabetic self-care practices?
- 2) What are the prevailing methodologies and techniques commonly utilized in the realm of diabetic self-care management to facilitate behavior modification?
- 3) Which theoretical frameworks and models can be effectively employed to underpin and guide the progress and application of diabetes self-care management applications?
- 4) What common aspects do diabetic self-care management programmes use today to effectively and completely treat the disease and empower patients?
- 5) What are the intricate challenges encountered in the current landscape of diabetes self-care applications, and what are the anticipated future directions and potential advancements?

In order to answer this question, the research builds following research objectives: i) to evaluate the extent to which existing diabetes management applications effectively support and facilitate self-care practices among individuals with diabetes. ii) to identify and analyze the methodologies and techniques commonly employed within the domain of diabetic self-care management. This includes an examination of strategies for behaviour modification and lifestyle improvement, iii) to explore and assess the theoretical frameworks and models that can be leveraged to underpin and guide the development and implementation of diabetes self-care management applications. iv) to identify the common elements and best practices employed by successful diabetic self-care management programs to comprehensively address the disease and empower patients in their self-care journey, v) to investigate the intricate challenges and limitations present in the current landscape of diabetes self-care applications. It will also explore anticipated future directions and potential advancements in the field to enhance the efficacy and usability of these applications.

The key significance of this study is that it underscores the importance of addressing various facets of diabetes management beyond medication, including lifestyle modifications and monitoring. It also offers valuable insights into how psychological and behavioral principles can improve diabetes self-management. It serves as a roadmap for future researchers and application developers to develop a more effective tools and interventions for diabetes self-care.

The rest of the sections are given as follows: Section II provides the detailed investigation of literature works on diabetes self-management and healthcare behavioral models, diabetes management interventions. Section III briefs the review methodology that comprises of the selection of articles for the review process. Section IV details the search results and analysis, and Section V briefs the outcomes and findings of the research questions. Section VI provides the overall discussion and Section VII finally concludes the study.

## II. RELATED WORKS

### A. Diabetes Mellitus

A major wellness concern and pandemic disease with a high incidence in both emerging and industrialized nations, diabetes impacts people all over the globe [24], [25]. As per the World Health Organization, diabetes is a chronic disorder with a variety of causes. The characteristic of this illness is prolonged hyperglycemia with abnormalities in carbohydrate, lipid, and protein metabolism induced by impairments in insulin action, insulin synthesis, or both. Additionally, diabetes-related complications and mortality could cause serious economic and social consequences for people, families, enterprises, and society overall [26]. People and medical institutions all across the world are being plagued by this epidemic [27], [28]. Diabetes presents a variety of dangerous side effects, involving dysfunction or long-term damage, and also organ failure [29]. There are numerous causes that might lead to a chronic illness like diabetes, but the following are the most typical ones: the pancreas' insulin is not generated properly or the pancreas has been unable to produce enough insulin. Additionally, if left untreated, increased blood glucose levels, also known as hyperglycemia, raise the risk of long-term harm to a range of organs, including blood vessels and neurons. Some of the diabetes-related symptoms include increased urine and weight loss, weariness, and increased appetite and thirst [30], [31].

Type 1 diabetes affects 10-15% of all diabetics and can appear at any age, with the majority of cases occurring in those under 40. It can be triggered by a range of critical variables such as infections, diet, and toxins in those who are genetically predisposed. An investigation [32] found that people with diabetes who have Type 1 have life expectancies that are around twelve decades lower than those of the entire community. The second type of diabetes, often called non-diabetes, has become the most frequent. It is responsible for 85-90 per cent of all diabetic patients [33]. Diabetes that occurs later in life is referred to as "late-onset diabetes." Dietary changes, a regular fitness routine, and medications could be treated Type-2 diabetes effectively. The third type of diabetes is gestational diabetes. Diabetes develops during pregnancy, unless the pregnant women already have been diagnosed, as a result of enhanced glucose levels or insulin levels. According to the International Diabetes Foundation, around 16% of women giving birth in 2019 have DM during pregnancy, with GD accounting for 85.1% of the total [34].

### B. Self-Management of Related Activities

The expression "self-management" denotes the routine tasks or activities that an individual needs to carry out for managing or lessening the effects of disease on their wealth and wellness in order to avoid further illness [35]. Medication adherence, physical exercise, healthy diet, monitoring, good coping, risk reduction, and problem-solving all seem to be examples of diabetes self-management practises that are important for better preventive measures [36]. Patients' adherence to diabetes self-management differed, showing that a number of factors will influence self-management decision-making mechanisms, either as enablers or as obstacles [37]. The care of diabetes mellitus is crucial for reducing long-term

effects and enhancing the T2D patients' quality of life. As per the American Diabetes Association, diabetes self-management-based education has been a pillar for optimum diabetes care. One viewpoint is that the complexity of T2D management necessitates the usage of DSME. Patients are assigned a variety of responsibilities, including a keeping regular doctor's appointments, confirming prescription schedules, and concentrating on self-care measures like online glucose tracking, healthy food modifications, and enhanced physical exercise [38]. However, people frequently struggle to maintain the many behavioral factors necessary for optimum glycemic control. Struggling to meet daily obligations, irritation, various types of mental discomfort, and an absence of self-commitment are all frequent problems [39]. Moreover, patients' levels of commitment to diabetic self-management varied, implying that many factors can affect self-management decision-making processes that could function as facilitators or barriers. DM management is critical for minimising long-term consequences and enhancing T2D patients' quality of life [36]. Therefore, consistent diabetic self-management was already linked to improve fewer complications [40], [41], blood glucose control, better quality of life, and a less peril of diabetes-related mortality [35]. Family members were able to offer both emotional and physical support. Instrument assistance could involve assisting patients with various chores, such as making appointments with healthcare providers or aiding with insulin treatment, and also assisting patients through self-management care. Providing delight and motivation to patients who've been frustrated or unhappy as a consequence of their therapeutic intervention is a common form of emotional support [8], [42]. Additional behavior change treatments are necessary because conventional techniques are ineffective at modifying behaviours [43]. In order to communicate with patients as well as give them the tools they need to manage their individual health; a web-based system is a suitable choice [13]. It has been demonstrated that diabetics' glucose levels could be improved through the implementation of mobile and internet-based diabetes care strategies [44].

### C. Concepts and Paradigms of Health Behavior's Impact on Diabetes Management Interventions

Theories of behavioural change aid in understanding human behaviour and change. They provide justification for why particular acts occur. These ideas are crucial for altering behaviour to improve health consequences [45]. Current years have seen the application of these theories to the management of persistent adherence to prescribed drugs and lifestyle changes [46]. Interventions for changing behaviour related to health may be more successful if they are based on the right hypothesis [47]. Through the identification of specific mediators—behaviour-causing variables, change-causing factors, and the mechanisms by which they operate during an intervention—theoretical models shed insight on basic concepts [48]. The following theoretical models are frequently used for strategizing and assessing public health behavioural change interventions: Health Belief Model (HBM), Transtheoretical Model, Social Cognitive Theory, Social Ecological Model, and Theory of Reasoned Action/Planned Behaviour [49], and Information-Motivation-Behavioural Skills models [50].

The Trans theoretical Model, also called as the Phases of Transformation Model [51], is one of the most well-known theories or models for health behaviour change that places a strong emphasis on the person's capacity for decision-making. Prochaska and DiClemente [52] created this model in the late 1970s based on research comparing the experiences of people who modify their conduct on their own to those who get therapy and how capable they are of doing so. This study's framework is based on this model. TTM underlines that people don't alter their behaviours right away but rather gradually, consistently, and through a cyclical process. There are five stages of change that individuals can go via in accordance with TTM: contemplation, planning, action, servicing, and relapsing [53]. When changing their lifestyle, everyone, even those with diabetes and prediabetes, typically goes through these stages. TTM provides comprehensive instructions on how to assist diabetics and prediabetics in making lifestyle and dietary changes that will promote healthy behaviour. The Theory of Reasoned Action (TRA) is expanded upon by the Theory of Planned Behaviour (TPB). Icek Ajzen put out this notion in 1985. According to TPB, an individual's willingness and degree of control over an activity impact how vigorously that behaviour will be carried out. An individual's behavioural intents and behaviour are influenced by their attitude towards a conduct, subjective standards, and perceived behavioural control [48], [54]. According to this principle, the concept of Personality is indirectly affected by beliefs that are influenced by background and demographic information like education, income, personality characteristics, prior behavior's, and aspects of the social and cultural environment [55].

The attitude towards behaviour, the importance others place on the behaviour, and the degree of perceived behavioural control all affect how strong an intention is. This applies to people with diabetes and prediabetes because in order to modify their conduct, they must also alter their mindset. They must also identify the triggers that encourage change [54]. The main significant factor influencing people's behaviour is their want to modify [54]. When patients want to change, they might alter their food and way of life. To forecast human social conduct, this theory is frequently applied and quoted [56]. Self-determination theory (SDT) encourages individuals to act in productive and beneficial ways. SDT emphasises a person's level of self-motivation and self-determination. Objectives and the pursuit of objectives are stressed in SDT. It suggests that what we are working for and why we are working towards it are both crucial for our wellbeing [57], [58]. According to the belief, if someone may pursue their objectives in their own way rather than being forced to adhere to rigorous rules, they would be happier and more successful. When someone pursues their goals for their own reasons and through their own ways, they would be happy and self-actualized [59]. By focusing on health-related advantages, patients will be more likely to accept personal responsibility for their health. The objectives for this must be independent and intrinsic. They should let to choose their own realistic goals. According to studies, SDT therapies for diabetes resulted in successful treatment outcomes [60]. It also encourages individuals by allowing them to identify the driving force behind transformation. Additionally, it is

observed that goal-setting is effective when supported by encouraging and compassionate individuals as opposed to dominating or directive people [61].

The Health Belief Model [62], [63], was created in the 1950s to address why certain individuals do not utilise the available health treatments. According to HBM, perceived vulnerability, severity, advantages, and obstacles all have an impact on conduct. The model describes and forecasts behaviour connected to health, including attitudes about one's health issues, the perceived advantages of taking action, obstacles, and self-efficacy in engaging in health-promoting activity. The health-promoting conduct should be triggered by cues to action. This approach was initially developed by social psychologists at the United States Department of Health and Human Services [64]. The concept of "perceived vulnerability" describes the probability that a person believes to be susceptible to get the illness if they continue with their current behaviors. On the contrary, perceived severity describes how serious the ailment is and how it affects people [65]. The perceived threat changes as a consequence of these behaviors. When people alter their behaviour, there could be apparent benefits or shortcomings of putting the unique activity into practise, among them perceived difficulties which could prevent the successful effectiveness, as well as both. For example, there might be perceived reduction in their chances of getting sick. The four factors mentioned above work together to affect the likelihood of participating in the behaviour. Information on the dangers and consequences of diabetes should be given to patients. Individuals must be made aware of the seriousness of the condition in order for them to change their bad behaviours and adopt healthy ones. Additionally, the advantages and drawbacks of their new outlook and behaviour ought to be explained to them. This should encourage individuals to adopt new habits and maintain their commitment to an improved diet and way of life [66].

Self-Regulation Theory describes the steps and elements involved in making decisions about one's thoughts, feelings, words, and actions. Self-regulation is concerned with the mechanisms that convert beliefs into intentions and intentions into actions, which ultimately results in the accomplishment of the connected objective. This idea focuses on a person's capacity to control their behaviours and their lives [67], [68]. SRT is made up of four components: requirements for desired behaviour, the drive to sustain standards, awareness of the conditions and thoughts that proceed standards-breaching, and willpower. SRT is centred on the idea of people setting the goals and monitoring their development in respect to those goals [69]. When there is a difference between their present situation and their aim during the comparison, individuals adjust their activities and behaviour in order to reach the goal. Diabetics should have the internal fortitude to alter their conduct in order to adjust their food and lifestyle. They should be self-motivated and devoted, with the purpose of transforming coming from inside. Self-control is crucial for developing new habits and viewpoints [70]. It aids patients in committing to their new conduct. In addition, students become more driven to accomplish their goals when they may define their own objectives and assess their progress in relation to

those objectives [71]. The Relapse Prevention Model seeks to impart knowledge on how to anticipate and address the issue of recurrence to persons wanting to change their behaviour's. Relapse occurs when a person fails to alter their conduct to match the desired behaviours. This approach proposes two methods for preventing relapses, which may be used either as a targeted maintenance plan or as a more comprehensive programme of lifestyle modification. The major goal of this paradigm is to change compulsive or addictive behavioural patterns.

One of the most well-liked theoretical paradigms for comprehending and altering health-related behavior's targeted at controlling persistent diseases is social cognitive theory (SCT). The SCT has proven significant behavioural changes leading to better health outcomes as the cornerstone of efficient illness self-management strategies [72], [73]. It began as the Social Learning Theory, which was referred to be the convergence of the cognitive and behaviourist approaches. Contrary to many other hypotheses of behavioral modification in health promotion, the concept of the SCT takes into consideration the unique ways that individuals develop and sustain a habit. In the most recent version of social cognitive theory, a complex causal structure is proposed in which beliefs about self-efficacy interact with knowledge of health hazards and advantages, targets, standards for the results, structural and social obstacles to modification, and the perceived facilitators of behavioral growth. In the temporal framework of the Social Comparative Theory self-efficacy plays a crucial regulatory function and is a fundamental belief that significantly impacts behaviour [74]. As per the Social-Cognitive Theory (SCT), interactions between the environment, a person's characteristics, and their behaviour affect behaviour change. The most significant influence on the acceptance of physical activity as a lifestyle change has come from the Social-Cognitive Theory [75]. It uses both cognitive and behavioural elements to encourage behaviour modification, comparable to the TTM [76]. Self-efficacy is the central construct of the social cognitive theory, but it also includes the concepts of social support, outcome expectancies, and self-regulation [77]. According to recent studies, while boosting a person's self-efficacy is vital for enhancing physical activity and exercise adherence, doing so is most successful when combined with using the other SCT elements. For the specialized maintenance approach, attention should be focused on strengthening the maintenance of behavior change, once a person has successfully predicted a behaviour change. The maintenance of behavioral change might take the shape of ongoing meetings, treatments, and other techniques that can make it last longer. Regarding the general one, the emphasis should be on facilitating variations in a person's habits and way of life. This general program's objectives are to instruct the client on how to live a balanced lifestyle and to stop the development of negative habit patterns [78].

The Information-Motivation-Behavioral Skills (IMBS) paradigm promotes the user-centered and evidence-based application of information in health-related situations [79], [79], [80]. It was first created to anticipate HIV preventive behaviour in response to the HIV epidemic. It was effectively used in the design of treatments that enhanced and predicted

adherence to medication among diabetic patients [81]. The IMBS offers a framework for comprehending and supporting disease prevention practises across populations, and it has a wide range of possible applications in health promotion practise [82]. The model focuses on a collection of components (factors) linked with illness management in terms of information, motivation, and behavioural skill. The model claims that behavioural changes are primarily brought about by changes in behaviour that occur as a consequence of informational and motivational interventions [80]. The third part of the model shows how knowledge, motivation, and the behavioural abilities needed to carry out self-management behaviours independently and to a substantial amount indirectly influence actions. When they regularly found a strong correlation among behavioural outcomes and IMBS and in cases of diabetes, academics and educators in diabetes health promotion have used IMBS. Information and motivation affect the behavioural skills of diabetes patients, ensuring that they have the resources necessary to engage in the desired actions. Finally, this boosts a patient's self-efficacy, or belief in their ability to carry out self-management actions [83]–[85].

#### D. Diabetes Prevention and Management Interventions

Diabetes is a chronic illness with a high incidence in many countries. It is marked by raised blood glucose levels and the possibility of both acute and chronic complications. It is generally recognized that treating diabetes is a difficult procedure that necessitates both a specific pharmacologic treatment plan and a change in lifestyle [86]. Effective behavioural change, thorough education, and self-management are some of the most important ways to prevent complications from diabetes. However, this procedure is time-consuming and costly. Recent research on the use of smartphone technology for managing diabetes has shown to be a useful tool for lowering haemoglobin levels, particularly in Type-2 diabetic (T2D) patients. The effectiveness of this approach among Saudi patients has not, however, been the subject of any recognised studies [87]. Diabetes management is a difficult procedure that needs a wide-ranging strategy. Pharmacologic therapy is crucial, but it must be supplemented with lifestyle changes such a nutritious diet, frequent exercise, and careful blood glucose monitoring. People with diabetes can effectively manage their illness lower their risk of complications, and lead satisfying lives by using these techniques. In order to support diabetic self-management (DSM), mobile phone applications are frequently utilised. Numerous apps have been created to improve diabetic self-management [87]–[120].

Numerous studies have found compelling proof that employing apps motivates individuals to stick to management medical care, enhances glycemic control, and delays or avoids the onset of diabetic complications while also improving their standard of life [10], [121]. Additionally, studies revealed that applications for diabetes self-care can dramatically enhance a number of clinical outcomes related to diabetes [5], [6]. Earlier research examined the effectiveness of self-care applications for diabetes qualitatively [7], [8]. Users' desires and requirements for self-empowerment applications, however, have changed over time. For example, prior to this,

the emphasis was primarily on the consumers administering their treatment alone with little help from the healthcare professionals and user preferences [15], [90], a large number of users, however, seem to anticipate that the applications would involve their healthcare providers in their regimens and routines, according to recent research [16], [122], [123]. The absence of customized feedback, poor user interface design, and accessibility concerns with current medical monitoring programmes are further problems (e.g. limited data entry options) [7], [15]–[17]. In addition, very few programmes have been created taking the needs of consumers into account [16], [85], [124]. As an outcome, many currently available programmes lack certain functionality [122]. It has been suggested that not enough thought has been given to end users' preferences as a cause of the low acceptance and utilisation of applications. Investigations are beginning in this area, and it is crucial to incorporate theories of health behaviour modification in the creation of diabetes management. Current research by Block et al. [125] stresses the benefits of the fully automated Alive-PD Diabetes Prevention Programme, which offers six to twelve months of weekly, step-by-step counselling on improving exercise, modifying eating habits, and losing weight . Although it is claimed that several applications have been created employing health behavioural change ideas, these theories have not yet been the subject of any study [126]. Additionally, there is additionally no appropriate outline for preventing diabetes that incorporates behavioral change theories and all other essential components [127].

### III. REVIEW METHODOLOGY

A PRISMA-based technique is used in the SLR approach. PRISMA provides a reliable and repeatable strategy for identifying literature. It also offers a manual for identifying, evaluating, and choosing research papers. In Fig. 1, the PRISMA procedure used in this SLR is depicted. The following subsections provide details on the SLR procedure:

#### A. Selection of Resources

The search process was carried out using nine digital online libraries to gather pertinent articles. In this study, Scopus, Google Scholar, ScienceDirect, Web of Science, SAGE, and Taylor & Francis Online were among the online databases that were investigated. These online databases were selected because they were thought to be the most ideal for offering comprehensive information in the area of older persons' social communities. While Scopus is a collection of peer-reviewed literature with approximately 22,000 articles from 5000 publishers worldwide, WoS is a powerful database with around 33,000 journals encompassing more than 250 subjects. In the area of diabetes treatment interventions, other digital libraries including Google Scholar, ScienceDirect, SAGE, and Taylor & Francis Online also have a sizable number of pertinent records available.

The articles from 2010 to 2022 were chosen for the SLR study in order to gain the most recent and complete review: (1) Appropriate resources pertaining to diabetes, mHealth apps (web-based and mobile apps), BCTs, and behavioural change theories make up the search phrase. Terms like "Mobile App" AND "Internet based Application" OR "diabetes" AND



"mHealth" OR ("behaviour change techniques" AND "diabetes" AND "mHealth" AND "Internet based application") are utilized to find more pertinent papers for this review. Moreover, the references of earlier literature were thoroughly examined in accordance with the comparable studies published in [9]–[11], [128], [129].

proceedings. Only conferences and journals are considered as types of literature; review articles, books, book series, and individual book chapters are not included. In order to clear up any ambiguity regarding translated literature, non-English publications were also expelled.

The chosen 36 articles were then imported into Zotero, a reference manager, for synthesis. The retrieved information includes information about the studies' and interventions' characteristics, such as the regions where the research studies took place, the platforms employed for the interventions, the percentage of baseline weight loss, and glycemic level. It should be pointed out that only the publicly accessible materials (such as the primary text, development procedures, supplementary materials, etc.) are taken into account for the extracting the data, identifying the application features, and BCTs coding processes because obtaining detailed information from authors remains challenging in many instances.

### C. Coding Scheme

To identify the existence or absence of every method from the evaluated articles, the list of BCTs taxonomy published in reference [130] was given special consideration. To more accurately evaluate the chosen research, the coding process was carried out individually and separately based on the primary publications, protocols, and related investigations. The initial research methods described in reference [131] and the BCT training materials were employed to create an accurate and suitable coding process for the BCT's taxonomy application. It was noted that equivalent research activities could have described the same standardised therapies based on earlier investigations in [132], [133]. But it was also noted that the interventions are described in a different way in the literature for every research, but with some BCTs having been identified in one study yet absent in another and vice versa. The present research uses an imputation approach to tackle these problems and recover the missing BCTs.

Based on the three phases of all intervention data, an analysis was done to find the characteristics in [134]. In order to accomplish this, every application component's description, coding, as well as the platform itself, must initially specified. The imputation process was also used in a scenario where numerous studies evaluated a similar standardised intervention. Furthermore, the characteristics were divided into two categories based on the degree of engagement between the user and the application: interactive (two-way interaction) and passive (one-way interaction). A random sample of all application descriptions was used for these first two steps in order to assess their dependability. Third, each interactive and passive component was collected, examined, and discussed as a whole. These findings led to the identification of common themes between the interactive and passive components. After being divided into interactive and passive elements, the themes or clusters were given individual labels.

### D. Quality Evaluation

To avoid the risk of biases in propagating a study, it is also essential to analyse the SLR data and assess its quality. Basically, an inadequately done study's outcomes could be

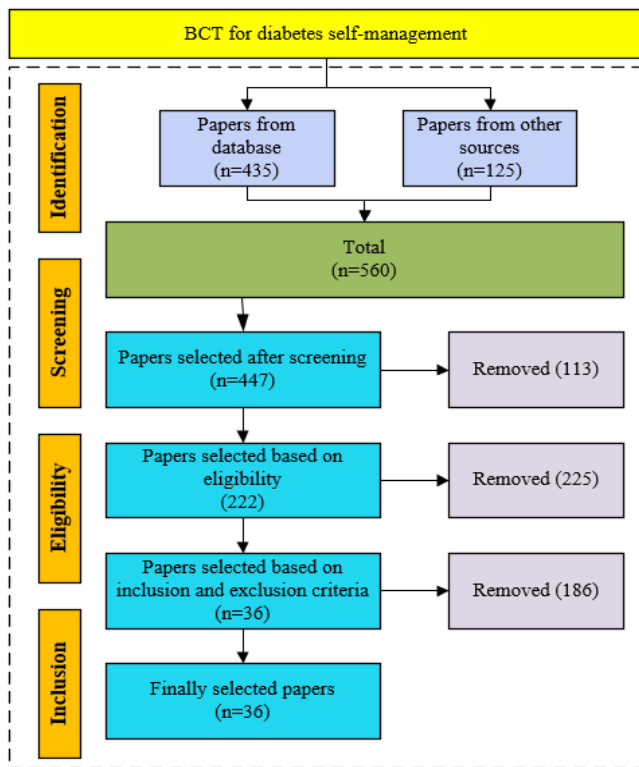


Fig. 1. The flow diagram of the review process based on PRISMA adopted in this study.

### B. Selection of Papers

A manual search was taken into account in addition to the search phrases used in the automatic search in order to completely recognize the pertinent research. The 500 research papers were found using keywords during the identification step, comprising 435 records found automatically when searching digital databases and 125 publications found manually by searching citations. After eliminating duplicates, inappropriate, and irrelevant publications, 447 published papers were chosen for the screening stage. 113 papers were ultimately eliminated after the remaining documents underwent additional screening based on titles and abstracts. After that, 222 more entries were subjected to the full-text evaluation. Following the removal of 186 articles based on the inclusion and exclusion criteria as well as the quality rating criteria, 36 articles remained. Regarding literature types, the effectiveness of self-care applications for adults 18 years and older at risk for getting diabetes was only investigated in article journals that concentrate on either research or design. For the current SLR, in particular, experimental, quasi-, and design investigations were taken into consideration. Many eligibility requirements and exclusion criteria are chosen. The studies that were chosen were those that have been peer-reviewed and published in English in journals or conference

greatly influenced by a considerable bias from the research process, necessitating careful interpretation. Consequently, in order to produce an objective result in the SLR, these studies must be disregarded or at the very least acknowledged as such. It is also crucial to evaluate the strength of the evidence and any inherent bias in every research investigation using the correct standards. For the quantitative intervention analysis, the National Institute of Health and Care Excellence (NICE) quality evaluation checklist is employed to validate the quality of the chosen studies in [135]. It includes 27 items that allow for the evaluation of external and internal validity when each criterion was met, with "++" denoting the lowest bias risk or greatest level of quality.

#### IV. SEARCH RESULTS AND ANALYSIS

##### A. Scholarly Publications over Time

Diabetes patients' self-management system seems to be a research study that will be crucial for society development in the future. In this part, the number of publications discovered over a fourteen-year period from 2010-2022 was selected. In Fig. 2, it shows how the quantity of papers has decreased during the last three years. The number of articles published starts from one in 2010, gradually increased to 4 in 2013, 5 in 2016, peaked to 7 in 2019 and afterwards rapidly declined to 1 in 2020 and increased to 3 in 2022.

##### B. Research Methods and Methodologies

Researchers employed a number of methodologies, including mixed method analysis, non-randomized controlled observational study, randomized controlled observational study, single arm prospective study and quasi-experimental methods to analyse the data connected to online supporting systems for diabetes self-management. The majority of these researches were predicated based on randomized control study. Both qualitative and quantitative methodologies were used in combination to support each other in a certain study. The distribution of included papers across research approaches is shown in Fig. 3.

As shown in Fig. 3, one study employed a quasi-experimental single arm technique. Furthermore, 4 employed

both observational study and non-randomized controlled observational study, 3 used single-arm prospective study, 6 employed mixed method design study, 2 employed prospective quasi-experimental study, and majority of the study, 15 articles employed randomized controlled trial-based study.

##### C. Publication Regions

The articles in this review came from all over the world, namely US, China, Australia, India, Saudi Arabia, Norway, Netherlands, Malaysia, Germany, Finland, Iran, Indonesia, Switzerland, Denmark, Italy and Sweden. In Fig. 4, it shows that the majority of the selected articles, 13 articles (36%) met our criteria from the US followed by China with 4% and Australia with 3%, India, Saudi Arabia and Norway with 2% each, and Netherlands, Malaysia, Germany, Finland, Iran, Indonesia, Switzerland, Denmark, Italy and Sweden with 1% each respectively.

The finding indicated that a large number of publications have been done for countries in the USA. The investigation was carried out in the Middle East, with a focus on Saudi Arabia. Furthermore, according to a WHO report, numerous individuals in Saudi Arabia have DM.

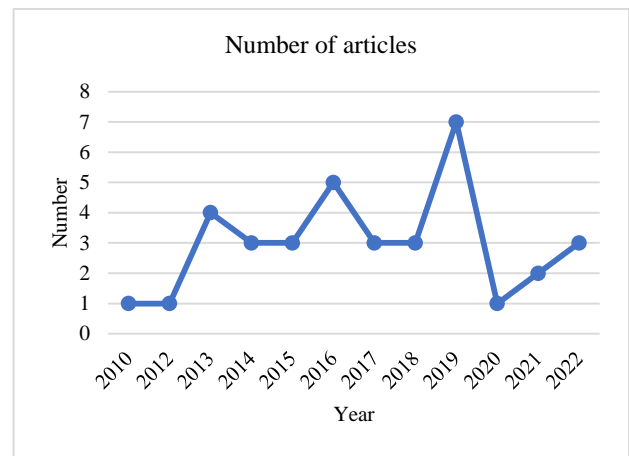


Fig. 2. Total amount of publications based on year.

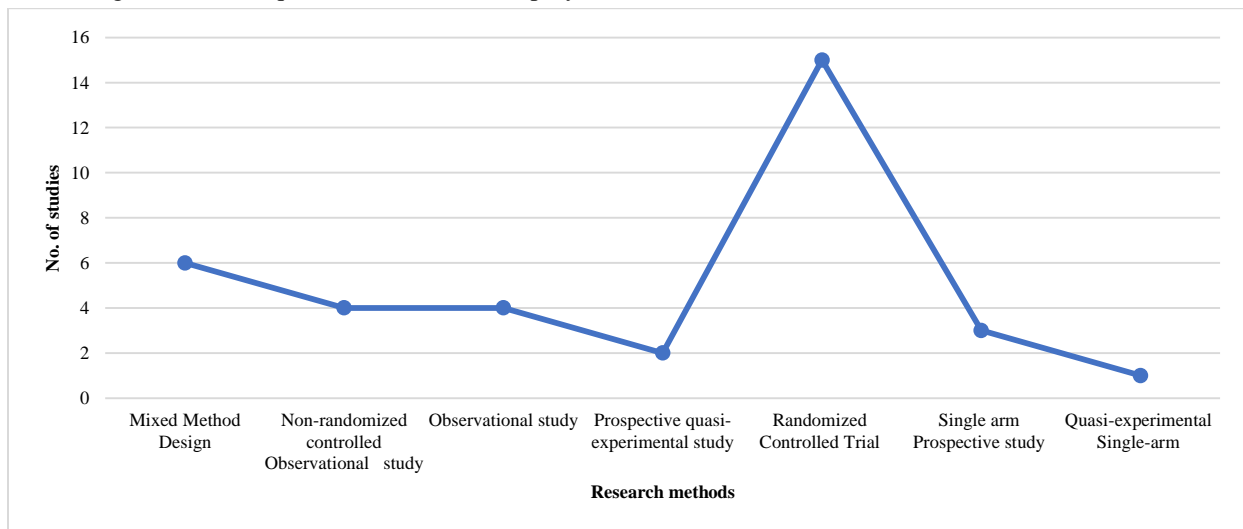


Fig. 3. Studies included over research approaches.

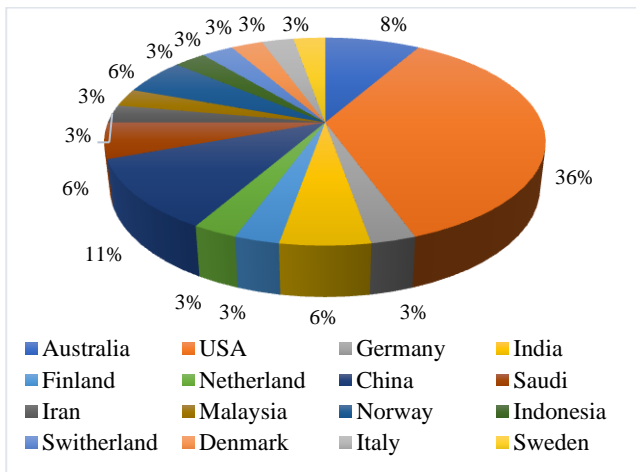


Fig. 4. Publication region/country.

#### D. Intervention Durations

Intervention duration in the reviewed articles ranged between 3 months and 24 months is showed in Fig. 5. More number of studies, 12 articles took approximately 3 months intervention duration, 10 articles took appropriately 6 months intervention duration, 6 articles took appropriately 12 months intervention duration, while rest of the 7 articles took more than 12 months of intervention duration.

#### E. Scholarly Articles Based on Theories

Fig. 6 shows the number of studies that have employed theory or model, ranging from Social Cognitive Theory (SCT) and Transtheoretical Model of Behavior (TTM) to Theory of Planned Behavior (TPB) and Health Belief Model (HBM). Other theories and models such as Self-Regulation Theory (SRT), Fogg Behavior Model, Cognitive Behavioral Therapy, COM-B model, IMB (Information-Motivation-Behavioral Skills Model), SDT (Self-Determination Theory), Just-in-time Adaptive intervention design, and Socio-material perspective have also been included.

From the Fig. 6, it is observed that the Social Cognitive Theory (SCT), Transtheoretical Model of Behavior (TTM), and Theory of Planned Behavior (TPB) are the most frequently used theories/models among the studies, whereas some theories/models like Fogg Behavior Model, Cognitive Behavioral Therapy, SDT (Self-Determination Theory), Just-in-time Adaptive intervention design, Socio-material and self-efficacy model have been utilized in a limited number of studies.

#### F. Scholarly Publications Based on Different Platform

Fig. 7 shows the various platforms used in various studies, showcasing the diverse approaches in delivering interventions or conducting research. Mobile apps emerged as the most frequently utilized platform, accounting for 56% of the studies. The combination of mobile app and web app platforms was employed in 19% of the studies, highlighting the recognition of multiple platforms' advantages. A smaller proportion of studies relied on websites (14%), DVDs (5%), and a combination of Short Message Service (SMS) and Email and a combination of website and mobile app (3%).

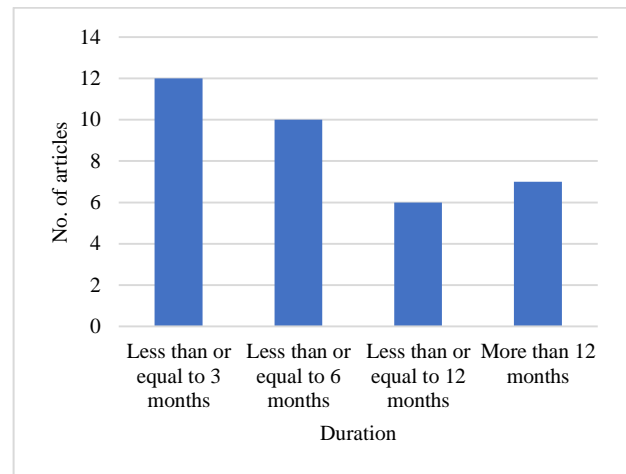


Fig. 5. Intervention durations.

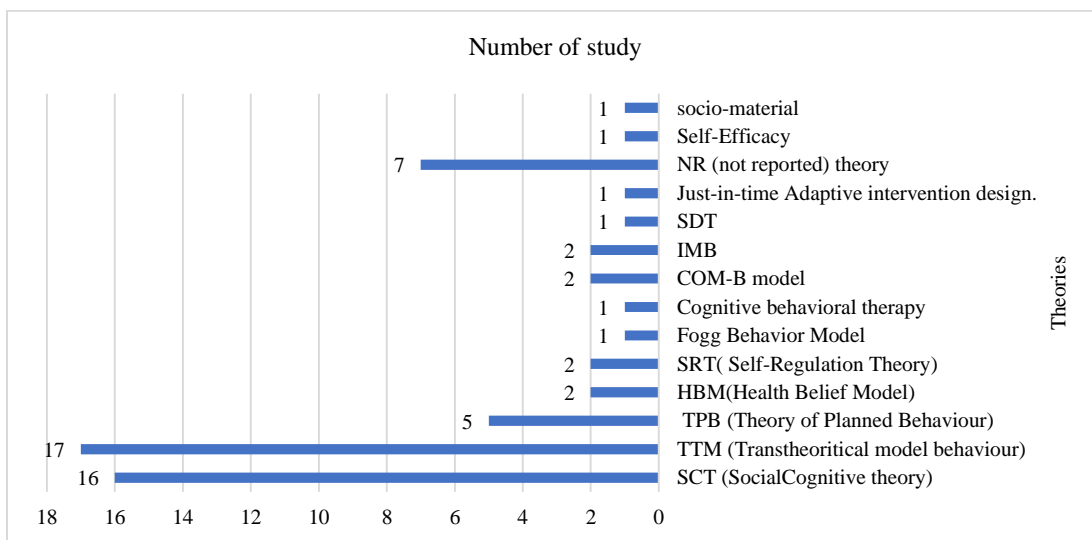


Fig. 6. Number of articles based on different theories.

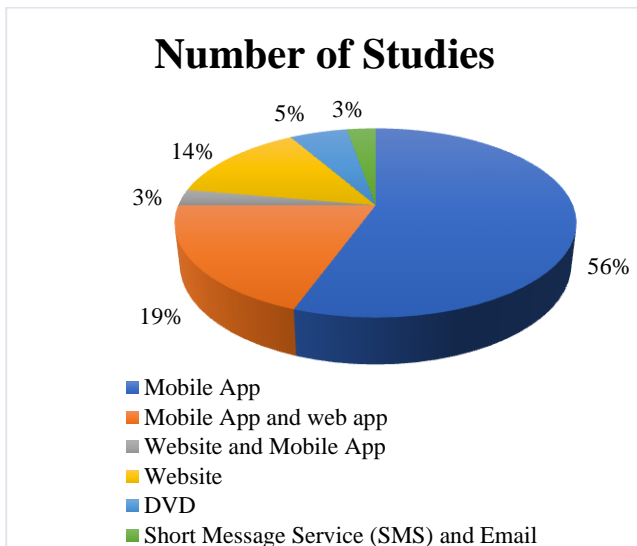


Fig. 7. Number of articles based on different platforms.

## V. RESEARCH QUESTION OUTCOME

*RQ1- How well does diabetes management applications support and facilitate diabetic self-care practices?*

In general, mobile diabetes interventions have shown promise in improving short-term outcomes for individuals with diabetes [87], [88], [90]–[92], [96], [98], [99], [101], [114], [115], [124], [136]–[138]. These interventions typically involve the use of mobile applications (apps) or other digital tools to help individuals manage their condition, track their blood glucose levels, monitor physical activity, and provide educational resources. Several studies have evaluated the effectiveness of mobile diabetes interventions over short-term periods (e.g., a few months to a year), and many have reported positive results that is short-term effectiveness [87], [88], [90]–[92], [96]–[98], [100], [101], [109], [114], [115], [124], [131], [136]–[138]. These interventions have been shown to improve glycemic control, increase self-management behaviors, enhance medication adherence, and promote healthy lifestyle choices, however, twelve interventions were short-term ineffective [91], [94], [98], [104], [105], [111], [113], [117], [139], [140]. Six interventions were long-term effective [92], [98], [110], [111], [136], [137]. Finally, ten interventions were long-term ineffective [101]–[106], [113], [114], [118], [139]. For example, some studies have found that mobile apps with features like glucose monitoring, medication reminders, and dietary guidance can lead to improvements in HbA1c levels (a measure of long-term blood glucose control) in the short term. Additionally, mobile interventions that include real-time feedback, coaching, and personalized recommendations have shown effectiveness in motivating individuals to adopt healthier behaviors. However, it's worth noting that the long-term effectiveness of these mobile interventions may vary. Some studies have reported challenges in maintaining the positive effects over a longer duration. Factors such as user engagement, adherence to the intervention, and sustainability of behavior change can influence the long-term effectiveness of these interventions. As technology advances and more research are conducted, it's

possible that newer mobile diabetes interventions may have improved long-term outcomes. It's essential for researchers and developers to continue evaluating the effectiveness and sustainability of these interventions to ensure their long-term benefits for individuals with diabetes.

*RQ2- What are the prevailing methodologies and techniques commonly utilized in the realm of diabetic self-care management to facilitate behavior modification?*

The term "behavioural change techniques" (BCTs) refers to discrete, observable, and repeatable elements of interventions intended to influence behaviour [131]. BCTs are a part of an intervention meant to change or restructure the causal mechanisms that control behaviour. The BCT Taxonomy, developed by Michie et al. in 2013, is a classification system for 93 hierarchically clustered approaches. Behavior change techniques are specific strategies or methods used to facilitate behavior change in individuals. These techniques are often employed in various fields, including healthcare, psychology, and public health, to promote positive behavior changes and support individuals in achieving their goals. BCTs can be used to modify a wide range of behaviors, including health behaviors like smoking cessation, physical activity, medication adherence, and dietary changes. They can be applied in individual counseling sessions, group interventions, digital health programs, or self-help materials. Behavior change techniques are designed to target specific determinants of behavior, such as motivation, self-efficacy, knowledge, and environmental factors. They are evidence-based and grounded in theories of behavior change, such as the Transtheoretical Model, Social Cognitive Theory, and the Theory of Planned Behavior.

Over the course of all the interventions examined, a total of thirty separate behaviour change methods (BCTs) were discovered, average 11.6 BCTs per intervention. Ten of these behaviour modification strategies were used in at least 55 per cent of both short and long-term interventions. In particular, behavioural goal-setting was used in 58.33% of interventions and was acknowledged in 75% of cases for both the short- and long-term categories. In 61.11% of the therapies that were considered, problem-solving was present, and its success was rated as being 75% short-term and 100% long-term. Defining outcome-related goals was another method that was used in 61.11% of all interventions, with recognition rates of 85% and 68.75% for short-term and long-term efficacy, respectively. Feedback on behaviours was noted in 61.11% of all interventions, with the rate of recognition for short and long-term interventions being 85% and 68.75%, correspondingly. In 80.56% of interventions overall, 100% of short-term treatments and 81.25% of long-term interventions showed evidence of the self-monitoring of behaviour technique. In line with this, the method of self-monitoring behavioural outcomes was recognised in 100% of short-term interventions and 75% of long-term interventions, or 77.7% of all interventions. Additionally, the undefined type of social support was recognised in 61.11% of all interventions, 95% of short-term interventions, and 62% of long-term interventions. Also noted in 50%, 55.56%, and 58.33% of all treatments, respectively, were strategies including informing participants about the potential health effects of their actions, citing reliable sources,

and modifying environmental signals. As mentioned in (see Fig. 8), for short-term and long-term interventions, respectively, the recognition rates for these approaches were 65%, 80%, and 90%.



Fig. 8. BCT interventions (a) Goal and planning cluster, (b) Feedback and monitoring cluster, (c) Shaping knowledge, (d) Social support cluster.

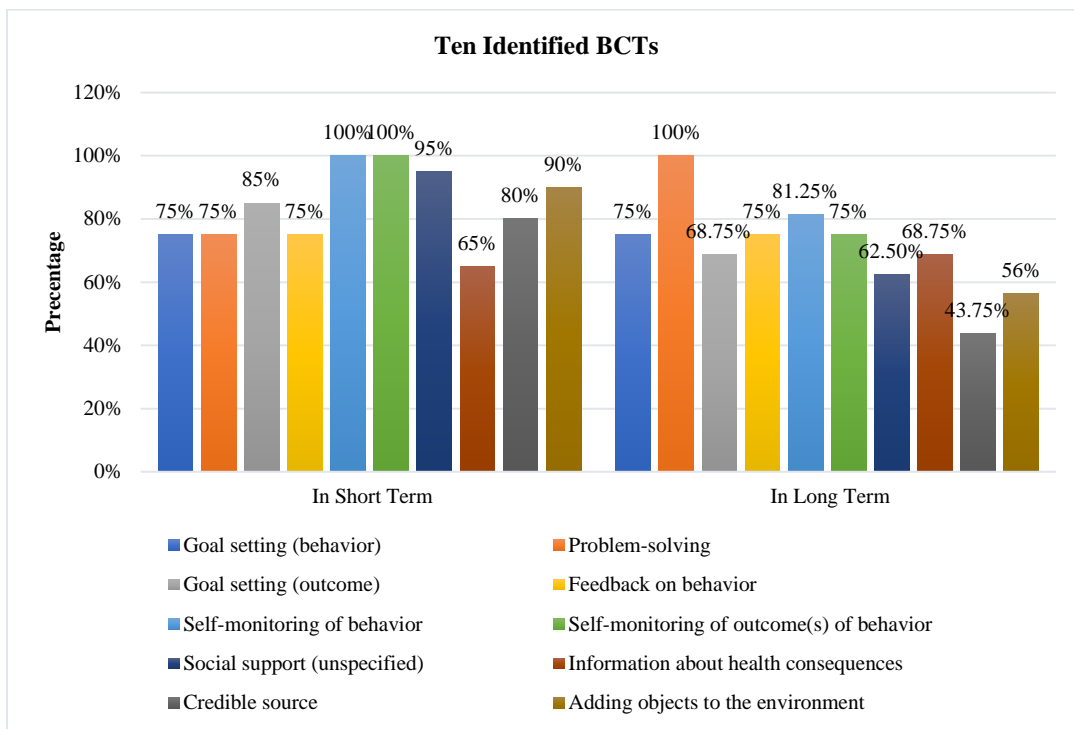


Fig. 9. Ten identified BCTs.

Fig. 9 depicts that the Goal setting (behavior) and problem-solving techniques are consistently employed in both short-term and long-term interventions, with percentages ranging from 75% to 100%. This suggests their recognized effectiveness in promoting behavior change. Goal setting (outcome) is more frequently utilized in the short term (85%), while its usage decreases slightly in the long term (68.75%), indicating a potential shift in focus over time. Feedback on behavior and self-monitoring of behavior are consistently utilized BCTs, highlighting their importance in promoting awareness and accountability. For both short-term and long-term therapies, the percentages range from 75% to 100%. Self-monitoring of behavior's result(s) is heavily used in the short term (100%) but less so in the long term (75%), suggesting a possible change in focus across various stages of behaviour change. Short-term (95%) but long-term (62.50%) use of social support (unspecified) declines, implying a potential shift to more focused types of social support with time. There are differences in percentages between short-term and long-term treatments when it comes to the use of information on health effects, reliable sources, and adding things to the environment. This implies that according to the particular environment and intervention goals, their efficacy and significance may change.

1) *Effectiveness in short term:* In comparison to long-term therapies, which used an average of 7.8 BCTs per intervention (range from 1 to 16), short-term interventions used an overall of 19 BCTs each intervention (range from 0 to 20). Self-monitoring of behaviours and self-monitoring of the results of behaviours were two behavioural change theories that were identified significantly more frequently in short-term therapies.

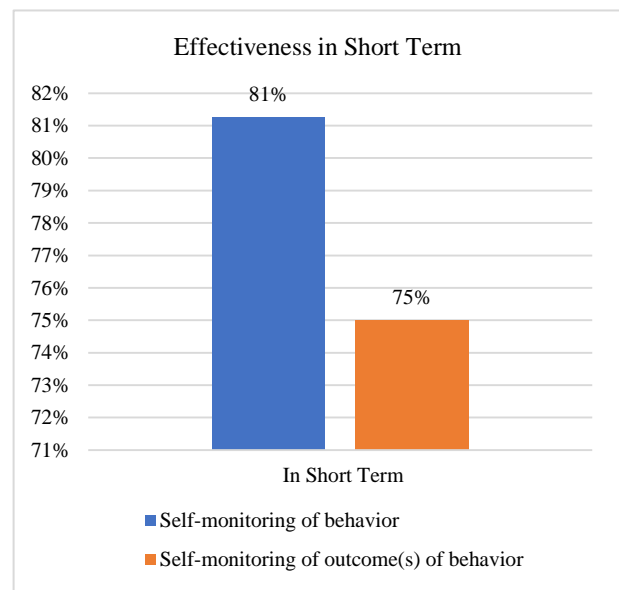


Fig. 10. Effectiveness in short-term.

As depicted in Fig. 10, the effectiveness of this BCT in the short term in self-monitoring of behavior is reported to be 81%. This suggests that when individuals actively monitor and track their behaviors, they are more likely to engage in positive self-care practices. While, the effectiveness of this BCT in the short-term self-monitoring outcome of behaviour is reported to be 75%, which implies that when individuals regularly track and observe the outcomes of their self-care behaviors, they can better understand the impact of their actions and make adjustments as needed.

2) *Effectiveness in long term:* In contrast to the 19 BCTs per intervention (range from 0 to 20) needed to achieve short-term efficacy, interventions that were long-term effective employed an average of 7.8 BCTs (ranging from 1 to 16). Two BCT found with noticeable higher frequency include action planning with 93.75%, and information about antecedents with 87.5%. As depicted in Fig. 11, the 87.50% effectiveness suggests that providing information about antecedents can be beneficial in promoting sustained behavior change and long-term self-care management. While, the 93.75% effectiveness indicates that when individuals engage in detailed action planning, they are more likely to maintain consistent self-care behaviors over an extended period.

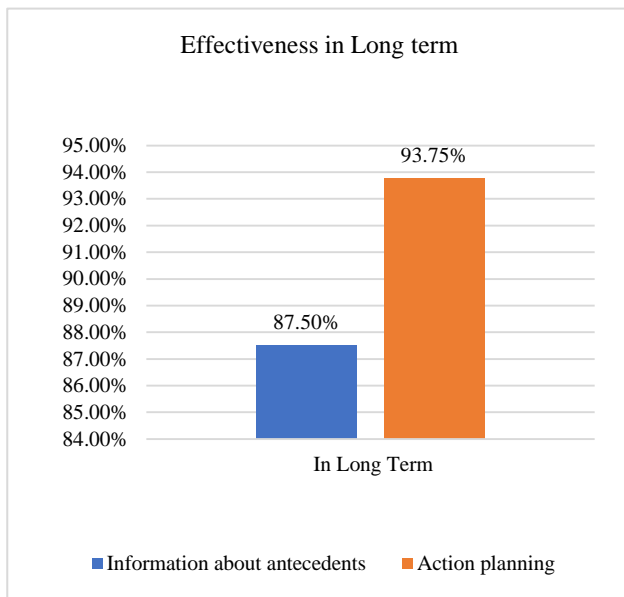


Fig. 11. Effectiveness in long-term.

*RQ3: Which theoretical frameworks and models can be effectively employed to underpin and guide the progress and application of diabetes self-care management applications?*

Multiple theoretical frameworks are frequently used to develop and assess public health interventions that aim to modify behaviors. The Health Behaviour Model examines a person's impression of a health problem's severity and vulnerability, in addition to the perceived advantages and challenges of implementing preventative behaviors. The TTM suggests that when changing behaviours, people go through phases of transformation. It emphasizes how crucial it is to adjust therapies depending to a person's level of progress. The Theory of Reasoning asserts Action/Planned Behaviors (TPB), a person's attitudes, personal standards, and perceived behavioral control all have an impact on their decision to engage in a behavior. It places a strong emphasis on how societal pressures and a person's beliefs might affect their behaviour. According to the Social Cognitive Theory (SCT), a person's behaviour, personal characteristics, and external factors all interact in a reciprocal manner. It emphasizes the significance of self-regulation and observational learning in behavior change. It also recognises that varieties of variables

such as those at the individual, interpersonal, communal, and societal levels, have an impact on behaviors. It highlights how these levels interact and the necessity of therapies that focus on several levels at once.

An overall of 29 interventions made reference to a theoretical underpinning for their design, whereas the other seven interventions made no such mention. Different behavioural change theories were used in the reviewed articles, which include Social Cognitive Theory (SCT) [141], [142], Theory of Planned Behaviour [143], [144], Transtheoretical Model (TTM) [145], [146] [147], Self-determination Theory [148], Information-Motivation-Behavioral Skills Model [142], Health Belief Model [144], [149]. Furthermore, COM-B model [150], Just-in-time Adaptive intervention design [151], Fogg Behavior Model [146], Self-efficacy [152] have been used in diabetes related interventions design. Twenty-nine interventions were supported (informed) by one theory or more theories. Some studies used several theories [142]–[144], [146], [148], [150], [153]–[166], while other interventions used a single behaviour theory [141], [151], [152], [165], [167]–[171].

Based on the reviewed articles, the most popular theories used in the studies were: Social Cognitive Theory (SCT) [141], [142], [153]–[157], [159]–[162], [164], [166], [172], [173] and the Transtheoretical Model of Behaviour Change (TTM) [144], [146], [154]–[157], [159], [160], [162], [163], [166], [169], [172]–[175]. SCT, which theorises knowledge gaining through social awareness and considers self-efficacy as one of the main channels of goal actualization, was used in 16 articles. TTM, which emphasizes changes as the progressive venture through pre-contemplating of behaviour change to behaviour maintenance [144], was informed in 17 articles. The Fig. 6 illustrate the distribution of the health behavioural theories in the reviewed articles.

According to Webb et al. [176] and Van Rhoon et al. [177], the Social Cognitive Theory and the TTM were among the most frequently utilised conceptual bases. On the other hand, the Theory of Planned Behaviour (TPB) represented one of the more often cited frameworks of theory in Webb et al.'s study [176]. Chao, Lin, and Ma [178] and Kusananto et al. [179] utilised concept only as an evaluation metric, while those that used concept as a component of the research design provided just a cursory description of how treatments were incorporated into the relevant theory.

Future study should focus on this issue, perhaps highlighting the importance of building a theoretical knowledge of the probable procedure for eliciting behaviour change at the outset of the conceptualization of an approach in [180] and, using Michie and Prestwish's description of the 'standardized' contribution of concept in [181]. Researchers in the future ought to be able to assess the efficiency of the role of concept in these kinds of interventions as well as possibly the relationship among the amount of theory utilised and the changes in behaviour and corresponding health outcomes. This could be possible with an explicit, systematic, and relatively consistent overview of the part of concept in the planning and creation of the intervention.

The management of sicknesses and overall well-being is clearly aided by self-care applications created utilizing health behavioural change techniques. These programmes are effective at encouraging patients to better adhere to their prescription regimens, encourage self-care, enhance their health, and lessen their despair. A number of investigations were undertaken to assess the accessibility of the programmes, and it was determined that these self-care apps were mainly simple to comprehend and utilize since the patients felt comfortable utilizing them and completing the necessary chores. Additionally, self-care behaviours and prevalent concepts employed in their creation have been identified. These concepts encompass the following: the Theory of Planned Behaviour, the model of health beliefs, Cognitive Behaviour Therapy, Self-Care Behaviour, Motivational Interviewing, and behaviour changes. Finding the health behaviour change paradigm that has been employed in previous research more frequently would be fascinating. According to the results of the investigations, every theory has been applied continually, whether solely or in conjunction with other approaches.

*RQ4- What common aspects do diabetic self-care management programmes use today to effectively and completely treat the disease and empower patients?*

Thematic analysis [182], was executed across three phases to uncover patterns within all interpolated data points. Initially, comprehensive explanations and codes were provided for every application components and its corresponding platform. Additionally, in cases where multiple studies evaluated identical standardized interventions, the imputation process was carried out. Subsequently, characteristics were classified according to the degree of engagement between the user and the application, categorized as either interactive (involving two-way interaction) or passive (involving one-way interaction). To test for dependability, we finished the initial two phases on an instance of randomly selected app specifications. Furthermore, all both active and passive characteristics were gathered, analyzed, and debated jointly. These findings led to the identification of common themes between the interactive and passive components. The interactive or passive characteristics of the themes or clusters were afterwards classified and labelled in accordance with each theme. In particular, several kinds of mobile and web-based therapies have tools for monitoring blood sugar, diet calories, and body weight, as well as alerts for remembering to take medications or schedule medical appointments. Consider dividing the characteristics into interactive and passive ones.

3) Digital feature descriptions: The different digital passive features are utilized in health interventions. Health and lifestyle information and advice provide educational materials on topics like healthy eating, physical activity, and stress reduction. Activity tracking involves tools like pedometers and accelerometers to record physical activity, while reminders and prompts send notifications to remind participants of specific tasks. Diet tracking allows participants

to record their dietary behaviors, including calorie counting and food diaries. Weight and bio-measure tracking involve tools like digital scales and blood glucose monitors to track body weight and biological measures. These features offer one-way interaction without active feedback.

Previous assessments of self-care apps have examined how the effectiveness of these applications is linked to their attributes in managing diabetes. Applications that demonstrated substantial effectiveness incorporated both passive and interactive features, while those with less pronounced effects tended to rely solely on passive attributes [183]. Passive attributes don't require user interaction, whereas interactive attributes involve real-time user engagement. In another investigation [184], diverse attributes in self-care apps were explored. This study revealed that interactive attributes were notably more successful than passive ones in enhancing medication adherence among individuals with type 2 diabetes. Nonetheless, this study exclusively concentrated on type 2 diabetes management, leaving uncertain the most efficacious features across various applications. Interactive elements encourage engagement and active involvement in programmes for a healthy lifestyle. The data shows that gamification, automatic feedback, social media and support, online medical coaching, and educating about healthy living all incorporate interactive components. Passive features include tracking nutrition, weight and measurements, suggestions and alerts, information on wellness and lifestyle, and activity monitoring. These elements are essential for providing people with information and insights so they may choose their lifestyle and health with knowledge. For instance, the majority of interventions incorporate health and lifestyle information, demonstrating the value of this information in educating people about several facets of their well-being [183].

The passive features include activity tracking, weight, biometric measurements, diet tracking, reminders and prompts and health and style information. From Fig. 12(a), it is observed that Health and lifestyle information is a commonly included passive feature in both short-term and long-term interventions. It is present in 58.33% of all interventions, 65% of short-term interventions, and 81.25% of long-term interventions. Activity tracking is another frequently incorporated passive feature, with 52.78% of all interventions including it. In the short term, all short-term interventions (100%) utilize activity tracking, while it decreases to 43.75% in the long term. Reminders and prompts are included in 47.22% of all interventions. They are used in 70% of short-term interventions and 56.25% of long-term interventions. Diet tracking is present in 41.67% of all interventions. It is used in 80% of short-term interventions but decreases significantly to 12.5% in the long term. Weight and measure tracking is utilized in 44.44% of all interventions. In the short term, 75% of interventions include this feature, but it drops to 6.25% in the long term. On average, short-term interventions have a higher number of passive features per intervention (3.81) compared to long-term interventions (2).



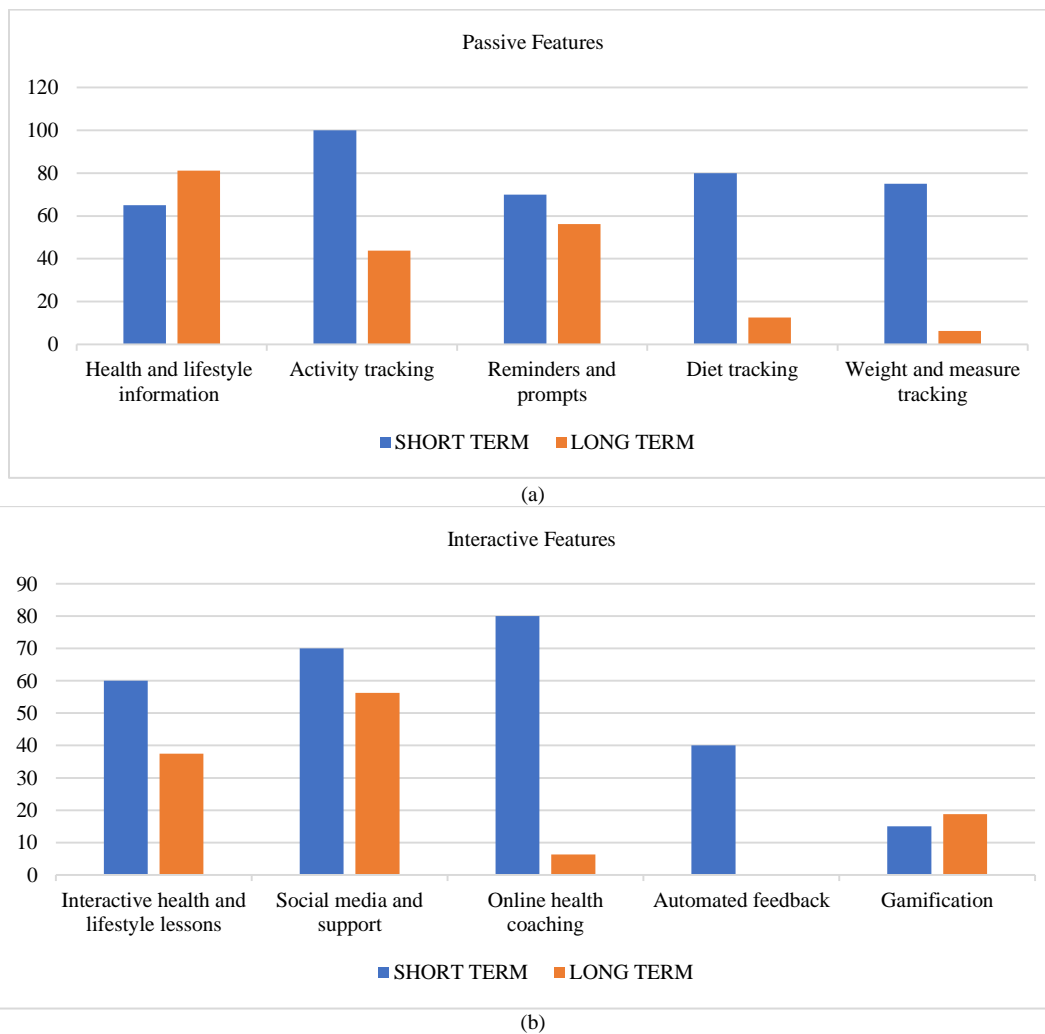


Fig. 12. (a) Passive features intervention and (b) Interactive features intervention.

The capabilities that are interactive include social media and support, automated feedback, lifestyle lessons, gamification, interactive wellness and online health instruction, and social networking sites. From Fig. 12(b), it is observed that Interactive health and lifestyle lessons are a commonly included interactive feature, present in 55.56% of all interventions. In the short term, 60% of interventions incorporate this feature, while in the long term, it decreases to 37.5%. Social media and support is another frequently included interactive feature, with 52.78% of all interventions including it. In the short term, 70% of interventions utilize social media and support, and in the long term, it remains high at 56.25%. Online health coaching is present in 50% of all interventions. It is used in 80% of short-term interventions but decreases significantly to 6.25% in the long term. Automated feedback is utilized in 36.11% of all interventions. In the short term, 40% of interventions include automated feedback, but it is not present in any of the long-term interventions. Gamification is the least frequently incorporated interactive feature, present in only 5.56% of all interventions. It is used in 15% of short-term interventions and 18.75% of long-term interventions. On average, short-term interventions have a

higher number of interactive features per intervention (2.52) compared to long-term interventions (1).

*RQ5: What are the intricate challenges encountered in the current landscape of diabetes self-care applications, and what are the anticipated future directions and potential advancements?*

While many diabetic self-management software programmes offer features like tracking diet and exercise, it's noteworthy that type 2 diabetes is frequently associated with insufficient dietary intake and insufficient physical activity. The features of machine learning (ML) or AI-powered dietary recommendation and planning, clinical support, fitness tracking, visualizing blood pressure, calorie expenditure estimation, and behavioural intervention (BI) techniques, still clearly have shortcomings. Additionally, most of these apps lacked well-recognized research underpinnings and ideas, including the nudge theory, which would add credibility. As a result, some diabetic self-management software programmes may not successfully help patients manage their condition on their own. It's interesting to note that none of these applications includes a food monitoring system powered by ML or AI, a system for providing individualised nutrition

advice, or a platform for meal recording (micronutrient detection) powered by AI using picture analysis. For diabetics to avoid hypo- or hyperglycemic episodes, improved nutrition management systems can significantly improve glycemic control [128]. These programmes, however, lack thorough feedback systems, such as organised behavioural agreements, consistent self-monitoring tracking, and goal setting. Based on the system's learned insights from recent data and preprogrammed guidelines, this input is customised for patients. Additionally, the majority of applications do not have AI-driven tools like insulin dose calculators that are intended to help patients make informed decisions by offering advice on activities, diets, and medications.

In future, a comprehensive self-care application for diabetes could be envisioned, encompassing both basic and advanced functionalities. This application would encompass elements like nutrition, blood glucose monitoring, clinical support, physical activity tracking, medication management, and tailored features. Moreover, this proposed diabetes self-care app would adopt an extensive feedback mechanism, fostering effective communication with all involved parties. Additionally, it would integrate behavioral intervention techniques guided by theories and artificial intelligence, promoting prolonged adherence to patients' self-care regimens.

## VI. DISCUSSION

A total of 36 therapies from multiple investigations were analysed and evaluated for this research. The SLR seeks to compile and evaluate the literature on diabetic self-care apps in order to assess the effectiveness of treatments for diabetes management. The Systematic Literature Review also looks for the best behavior change methods (BCTs) and application features that are frequently employed in the research investigations that are already out there. This study showed that, in the short term, a substantial number of diabetic self-care strategies resulted in noticeable weight loss, as evidenced by an average weight reduction of at least 3% from the original level, based on the evaluations that were chosen. However, after taking into account the longer time, the majority of therapies did not meet the clinically significant threshold of 5% weight loss. Previous reports [9]–[11], [128], [129] on the effectiveness of diabetes self-care applications found similar results and heterogeneity among studies. According to earlier research [121], applications that used more behaviour change methods (BCTs) typically had higher efficacy. Within the interventions, seven typical behaviour modification strategies were found, including goal formulation, self-monitoring, feedback on behaviour, problem-solving outcome, self-monitoring, and social support. The suggested behavior modification elements listed in the IMAGE toolkit for diabetes prevention are in line with these identified BCTs, which is significant [185]. The most successful behavior change method involves participants in problem-solving activities that encourage them to come up with potential behavioural modification methods, choose the best one, and implement it. According to the research, applications that included more behavior change methods were generally more likely to be more successful. Furthermore, consistent patterns of Behavior Change

Techniques (BCTs) were identified in both long-term and short-term therapeutic interventions. Notably, interventions with a higher number of distinctive attributes exhibited greater effectiveness, aligning with behavior change strategies. Similar findings were documented in studies [186] and [187], illustrating that self-care applications yielded enhanced efficacy in diabetes management through the incorporation of interactive features. Three elements have been frequently mentioned as useful interventions, indicating that they may make up an efficient core collection that would serve as the foundation for all subsequent applications.

In order to overcome the constraints to face-to-face interventions' connectivity, self-care applications have been developed. According to the present research, interventions that use more BCTs and characteristics are more effective, and because of their enhancing capabilities and adoption rates, websites and smartphones might serve as the best platforms for these strategies. Given that these behaviors are similarly comparable to the evidence-based treatment, that mostly depends on how the concepts are applied in the intervention layout, health behavioral theories have been shown to be an essential strategy for promoting behavioural modifications such as physical activity and nutritious eating [188]–[190]. While technology has the potential to positively impact self-care in diseases like diabetes, it alone is not enough. Effective outcomes rely on tailoring information appropriately and ensuring patients are highly motivated [191]. According to research, the broad implementation of BCTs, features, and theories into mobile and web-based therapies increases their effectiveness [188], [189]. Because there has been less prior research in this field, the creation of self-care interventions is an essential and developing direction in information science. However, current studies in the sector have shown that programmes created with the integration of BCTs and behavioural change theories produce better clinical results in [188], [189]. It may be concluded that in order to attain long-term effectiveness and help users reach their therapeutic goals, an ideal self-management intervention should incorporate BCTs and behavioural health theories during the design process.

Theoretical frameworks for influencing behaviors to improve medical results are crucial. Due to their ability to provide light on human behaviour and change, these theories are crucial components of successful intervention. By incorporating methods for behavioral change, these theories could be utilised in the creation of fresh apps. Researchers offer encouraging recommendations for developing, putting into practice, and assessing health promotion programmes that might be incorporated into the creation of self-care interventions dealing with health-related behaviours [187]. BCT-based therapies may be employed to encourage users to improve their health-related behaviours [131].

Any health support programmes that include more behavioral change theories are thought to be more successful at achieving the intended behavioral change. Many research investigations have looked into how health behavioral change theories could be included into the creation of medical assistance apps. However, few researches have focused on such incorporation in prediabetes self-care strategies.

According to the findings of the prediabetes research, it would be achieved to avoid diabetes and manage prediabetes—but only when prediabetics are inspired to take charge of their well-being by altering their attitudes through self-care behaviors. In the present research, examine the efficacy of existing theories of health behavior change as they are applied to prediabetes therapies worldwide and assess the efficacy of self-care apps that combine these concepts. The development of self-care strategies that target behaviours associated with health could be influenced by behavioural change concepts and approaches, which offer promising principles for creating, carrying out, and evaluating health promotion programmes [192]. Understandings how individuals act and modification could assist us achieve better health effects, according to behavioural change concepts [64]. Incorporating ideas of wellness behavior modification into the creation of healthcare applications has been the subject of several research. However, few researchers have focused on this inclusion in diabetic self-care programs [91], [93]. According to the findings of the mellitus research, managing and preventing diabetic is only feasible if individuals are inspired to take control of their well-being by altering their attitudes through self-care behaviors. As a result, researchers advise the investigators to create diabetes self-management software that includes both simple and sophisticated features, including dietary advice, fitness advice, calorie prediction, and insulin bolus calculations. The software should also facilitate stakeholder communication, incorporate theory based on artificial intelligence that improves the programme's effectiveness in managing diabetes, and allow diabetic patients to commit to their self-management regimens progressively over time.

As previously mentioned, this study's investigation of the effectiveness of self-care applications for managing diabetes and the effects of the BCTs and application features is one of its contributions. As a result, the information obtained from the examined papers was presented and organized using a narrative synthesis technique, with tables that summarised the descriptive analysis and statistical data. The data are more than adequate to perform a complete meta-analysis, but, the majority of the examined articles included in the primary efficiency analysis of the therapies did not present a percentage of weight loss and other essential requirements.

Furthermore, it was believed that body weight and glycemic status (A1c) were the main outcomes of importance. Due to its relationship to diabetic issues and the fact that diabetic self-care research commented on it more often than publications in other fields, body weight was seen as the major measure of success [193]–[195]. Since this value is considered to be clinically important [196] and usually complies with standards of weight loss for twelve-month diabetes self-care therapy [195], [197]. and the effectiveness of the intervention was assessed in terms of an average weight loss of fewer than five per cent from the starting weight.

In essence, interventions lasting less than six months were deemed successful if an average weight loss of over 3% occurred within this timeframe. For interventions extending beyond twelve months, success was determined if an average weight reduction of 5% or more was achieved within a twelve-

month follow-up period. Based on these criteria, the applications in the studies under review were categorized as either short-term effective, short-term ineffective, long-term effective, or long-term ineffective. Specifically, interventions exceeding twelve months were grouped into short-term (ST) and long-term (LT) follow-ups. The study investigated relationships between types of Behavior Change Techniques (BCTs) and intervention attributes identified in long-term compared to short-term interventions. Similar to pertinent findings in reference [196], effective BCTs and features were identified for each respective time (ST or LT) if present in a minimum of 55% of effective interventions.

## VII. CONCLUSION

This study focused on assessing the impact of self-care apps, particularly those integrating Behavioral Change Techniques (BCTs) and related concepts, in managing diabetes compared to standard treatment. The findings from the analysis of various studies suggest that the use of self-care apps can lead to improvements in A1c levels and weight loss for individuals with diabetes when compared to standard care. These results align with previous research, reinforcing the potential benefits of self-care apps in diabetes management. It also demonstrated that previous studies that utilised behavioral health concepts and BCTs in the creation of diabetic self-management treatments tended to be more successful. Importantly, the analysis highlights that the incorporation of BCTs and related concepts into interventions is associated with a reduction in A1c levels. This underscores the significance of integrating these strategies into self-care applications, even though the precise influence on application features can sometimes be unclear.

### A. Limitations and Future Directions

After determining the possibility for incorporating BCT theories and practises into the creation of self-management interventions, it is vital to build a paradigm for creating successful self-care programmes based on particular BCT. As a foundation for developing future applications, it is also necessary to explicitly elaborate on the use of BCTs and concept. While this systematic literature review (SLR) provides valuable insights, it's important to acknowledge its limitations. The studies included in this review varied in terms of methodology, participant characteristics, and intervention design, which may introduce heterogeneity into the findings. To build on these findings, future research should consider more standardized methodologies and explore the long-term effects of self-care apps. Furthermore, research should focus on elucidating the specific mechanisms through which BCTs and related concepts influence the efficacy of self-care interventions for diabetes management. Such efforts could provide a clearer foundation for the development of more effective self-care programs and applications in the future.

## ACKNOWLEDGMENT

The authors would like to express gratitude to Universiti Teknologi, Malaysia for the financial sponsorship of the research through the UTM Encouragement Research Grant (UTMER), grant reference number/no: PY/2022/03968; cost center: Q.J130000.3828.31J51

REFERENCES

- [1] S. Joachim, N. Wickramasinghe, P. P. Jayaraman, A. Forkan, and A. Morshed, "Design and development of a diabetes self-management platform: a case for responsible information system development," 2021.
- [2] Kusnanto, K. A. J. Widyanata, Suprajitno, and H. Arifin, "DM-calendar app as a diabetes self-management education on adult type 2 diabetes mellitus: a randomized controlled trial," *J Diabetes Metab Disord*, vol. 18, no. 2, pp. 557–563, Dec. 2019, doi: 10.1007/s40200-019-00468-1.
- [3] D. Tomic, J. E. Shaw, and D. J. Magliano, "The burden and risks of emerging complications of diabetes mellitus," *Nature Reviews Endocrinology*, vol. 18, no. 9, pp. 525–539, 2022.
- [4] M. M. McCarthy, R. Whitemore, G. Gholson, and M. Grey, "Diabetes Distress, Depressive Symptoms and Cardiovascular Health in Adults with Type 1 Diabetes," *Nursing research*, vol. 68, no. 6, p. 445, 2019.
- [5] X. Lin et al., "Global, regional, and national burden and trend of diabetes in 195 countries and territories: an analysis from 1990 to 2025," *Scientific reports*, vol. 10, no. 1, p. 14790, 2020.
- [6] H. Sun et al., "IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045," *Diabetes research and clinical practice*, vol. 183, p. 109119, 2022.
- [7] O. El-Gayar, P. Timsina, N. Nawar, and W. Eid, "Mobile applications for diabetes self-management: status and potential," *Journal of diabetes science and technology*, vol. 7, no. 1, pp. 247–262, 2013.
- [8] B. Hansel et al., "A Fully Automated Web-Based Program Improves Lifestyle Habits and HbA1c in Patients With Type 2 Diabetes and Abdominal Obesity: Randomized Trial of Patient E-Coaching Nutritional Support (The ANODE Study)," *J Med Internet Res*, vol. 19, no. 11, p. e360, Nov. 2017, doi: 10.2196/jmir.7947.
- [9] K. Liu, Z. Xie, C. K. Or, and others, "Effectiveness of mobile app-assisted self-care interventions for improving patient outcomes in type 2 diabetes and/or hypertension: systematic review and meta-analysis of randomized controlled trials," *JMIR mHealth and uHealth*, vol. 8, no. 8, p. e15779, 2020.
- [10] Y. Mao, W. Lin, J. Wen, and G. Chen, "Impact and efficacy of mobile health intervention in the management of diabetes and hypertension: a systematic review and meta-analysis," *BMJ Open Diabetes Research and Care*, vol. 8, no. 1, p. e001225, 2020.
- [11] X. Wu, X. Guo, and Z. Zhang, "The Efficacy of Mobile Phone Apps for Lifestyle Modification in Diabetes: Systematic Review and Meta-Analysis," *JMIR Mhealth Uhealth*, vol. 7, no. 1, p. e12297, Jan. 2019, doi: 10.2196/12297.
- [12] H. B. Aminuddin, N. Jiao, Y. Jiang, J. Hong, and W. Wang, "Effectiveness of smartphone-based self-management interventions on self-efficacy, self-care activities, health-related quality of life and clinical outcomes in patients with type 2 diabetes: A systematic review and meta-analysis," *International journal of nursing studies*, vol. 116, p. 103286, 2021.
- [13] M. Hood, R. Wilson, J. Corsica, L. Bradley, D. Chirinos, and A. Vivo, "What do we know about mobile applications for diabetes self-management? A review of reviews," *Journal of behavioral medicine*, vol. 39, pp. 981–994, 2016.
- [14] E. Gong et al., "Quality, functionality, and features of Chinese mobile apps for diabetes self-management: systematic search and evaluation of mobile apps," *JMIR mHealth and uHealth*, vol. 8, no. 4, p. e14836, 2020.
- [15] J. Pavlas, O. Krejcar, P. Maresova, and A. Selamat, "Prototypes of User Interfaces for Mobile Applications for Patients with Diabetes," *Computers*, vol. 8, no. 1, p. 1, Dec. 2018, doi: 10.3390/computers8010001.
- [16] R. H. Franklin, M. Waite, and C. Martin, "The use of mobile technology to facilitate self-management in adults with type 1 diabetes: A qualitative explorative approach," *Nursing open*, vol. 6, no. 3, pp. 1013–1021, 2019.
- [17] M. D. Adu, U. H. Malabu, A. E. O. Malau-Aduli, and B. S. Malau-Aduli, "The development of My Care Hub Mobile-Phone App to Support Self-Management in Australians with Type 1 or Type 2 Diabetes," *Sci Rep*, vol. 10, no. 1, p. 7, Jan. 2020, doi: 10.1038/s41598-019-56411-0.
- [18] S. Baptista, S. Trawley, F. Pouwer, B. Oldenburg, G. Wadley, and J. Speight, "What do adults with type 2 diabetes want from the 'perfect' app? Results from the second diabetes MILES: Australia (MILES-2) study," *Diabetes technology & therapeutics*, vol. 21, no. 7, pp. 393–399, 2019.
- [19] L. F. Garabedian, D. Ross-Degnan, R. F. LeCates, and J. F. Wharam, "Uptake and use of a diabetes management program with a mobile glucometer," *Primary Care Diabetes*, vol. 13, no. 6, pp. 549–555, 2019.
- [20] Y. Shen et al., "Effectiveness of internet-based interventions on glycemic control in patients with type 2 diabetes: meta-analysis of randomized controlled trials," *Journal of medical Internet research*, vol. 20, no. 5, p. e172, 2018.
- [21] L. E. Pathak et al., "Developing messaging content for a physical activity smartphone app tailored to low-income patients: user-centered design and crowdsourcing approach," *JMIR mHealth and uHealth*, vol. 9, no. 5, p. e21177, 2021.
- [22] L. A. Nelson, S. A. Mulvaney, K. B. Johnson, and C. Y. Osborn, "mHealth intervention elements and user characteristics determine utility: a mixed-methods analysis," *Diabetes Technology & Therapeutics*, vol. 19, no. 1, pp. 9–17, 2017.
- [23] V. N. Shah and S. K. Garg, "Managing diabetes in the digital age," *Clinical Diabetes and Endocrinology*, vol. 1, pp. 1–7, 2015.
- [24] M. A. Basar et al., "A review on diabetes patient lifestyle management using mobile application," in 2015 18th International Conference on Computer and Information Technology (ICCIT), IEEE, 2015, pp. 379–385.
- [25] R. Itumalla, R. Kumar, M. Tharwat Elabbasy, B. Perera, and M. R. Torabi, "Structural Factors and Quality of Diabetes Health Services in Hail, Saudi Arabia: A Cross-Sectional Study," in *Healthcare*, MDPI, 2021, p. 1691.
- [26] C. M. Marx, "Economic implications of type 2 diabetes management," *The American journal of managed care*, vol. 19, no. 8 Suppl, pp. S143–S148, 2013.
- [27] M. Aljofan, A. Altebainawi, and M. N. Alrashidi, "Public knowledge, attitude and practice toward diabetes mellitus in Hail region, Saudi Arabia," *International Journal of General Medicine*, pp. 255–262, 2019.
- [28] A. B. Kaiser, N. Zhang, and W. V. Der Pluijm, "Global prevalence of type 2 diabetes over the next ten years (2018-2028)," *Diabetes*, vol. 67, no. Supplement\_1, 2018.
- [29] U. Asmat, K. Abad, and K. Ismail, "Diabetes mellitus and oxidative stress—A concise review," *Saudi pharmaceutical journal*, vol. 24, no. 5, pp. 547–553, 2016.
- [30] L. Hernandez, H. Leutwyler, J. Cataldo, A. Kanaya, A. Swislocki, and C. Chesla, "The symptom experience of older adults with Type 2 diabetes and diabetes-related distress," *Nursing research*, vol. 68, no. 5, p. 374, 2019.
- [31] D. García-Huidobro, M. Bittner, P. Brahm, and K. Puschel, "Family intervention to control type 2 diabetes: a controlled clinical trial," *Family practice*, vol. 28, no. 1, pp. 4–11, 2011.
- [32] L. Huo, J. L. Harding, A. Peeters, J. E. Shaw, and D. J. Magliano, "Life expectancy of type 1 diabetic patients during 1997–2010: a national Australian registry-based cohort study," *Diabetologia*, vol. 59, pp. 1177–1185, 2016.
- [33] J. Apelqvist, "The diabetic foot syndrome today: a pandemic uprise," in *The Diabetic Foot Syndrome*, Karger Publishers, 2018, pp. 1–18.
- [34] A. A. Muche, O. O. Olayemi, and Y. K. Gete, "Prevalence and determinants of gestational diabetes mellitus in Africa based on the updated international diagnostic criteria: a systematic review and meta-analysis," *Archives of Public Health*, vol. 77, pp. 1–20, 2019.
- [35] S. SH and M.-C. Huang, "Diabetes Self-Management Engagement: A Case Study Analysis of Respect for Patient's Autonomy," 2020.
- [36] R. A. Pamungkas, K. Chamroonsawasdi, and P. Vatanasomboon, "A systematic review: family support integrated with diabetes self-management among uncontrolled type II diabetes mellitus patients," *Behavioral Sciences*, vol. 7, no. 3, p. 62, 2017.
- [37] Y. M. Alneami and C. L. Coleman, "Risk factors for and barriers to control type-2 diabetes among Saudi population," *Global journal of health science*, vol. 8, no. 9, p. 10, 2016.

- [38] L. B. Cohen, T. H. Taveira, S. A. M. Khatana, A. G. Dooley, P. A. Pirraglia, and W.-C. Wu, "Pharmacist-led shared medical appointments for multiple cardiovascular risk reduction in patients with type 2 diabetes," *The Diabetes Educator*, vol. 37, no. 6, pp. 801–812, 2011.
- [39] W. T. Tong, S. R. Vethakkan, and C. J. Ng, "Why do some people with type 2 diabetes who are using insulin have poor glycaemic control? A qualitative study," *BMJ open*, vol. 5, no. 1, p. e006407, 2015.
- [40] T. D. Anekwe and I. Rakhovsky, "Self-management: a comprehensive approach to management of chronic conditions," *American Journal of Public Health*, vol. 108, no. S6, pp. S430–S436, 2018.
- [41] P. A. Grady and L. L. Gough, "Self-management: a comprehensive approach to management of chronic conditions," *American journal of public health*, vol. 104, no. 8, pp. e25–e31, 2014.
- [42] A. M. AlHaidar, N. A. AlShehri, and M. A. AlHussaini, "Family Support and Its Association with Glycemic Control in Adolescents with Type 1 Diabetes Mellitus in Riyadh, Saudi Arabia," *Journal of Diabetes Research*, vol. 2020, pp. 1–6, Mar. 2020, doi: 10.1155/2020/5151604.
- [43] K. K. Berhe, H. B. Gebru, and H. B. Kahsay, "Effect of motivational interviewing intervention on HgbA1C and depression in people with type 2 diabetes mellitus (systematic review and meta-analysis)," *PloS one*, vol. 15, no. 10, p. e0240839, 2020.
- [44] X. Zhuo, P. Zhang, L. Barker, A. Albright, T. J. Thompson, and E. Gregg, "The lifetime cost of diabetes and its implications for diabetes prevention," *Diabetes care*, vol. 37, no. 9, pp. 2557–2564, 2014.
- [45] M. Conner and P. Norman, *EBOOK: predicting and changing health behaviour: research and practice with social cognition models*. McGraw-hill education (UK), 2015.
- [46] J. McSharry et al., "Behaviour change in diabetes: behavioural science advancements to support the use of theory," *Diabetic Medicine*, vol. 37, no. 3, pp. 455–463, 2020.
- [47] N. C. Campbell et al., "Designing and evaluating complex interventions to improve health care," *Bmj*, vol. 334, no. 7591, pp. 455–459, 2007.
- [48] W. Hardeman et al., "A causal modelling approach to the development of theory-based behaviour change programmes for trial evaluation," *Health education research*, vol. 20, no. 6, pp. 676–687, 2005.
- [49] S. E. Linke, C. J. Robinson, and D. Pekmezi, "Applying psychological theories to promote healthy lifestyles," *American Journal of Lifestyle Medicine*, vol. 8, no. 1, pp. 4–14, 2014.
- [50] E. L. Tuthill et al., "Exclusive breast-feeding promotion among HIV-infected women in South Africa: an Information–Motivation–Behavioural Skills model-based pilot intervention," *Public health nutrition*, vol. 20, no. 8, pp. 1481–1490, 2017.
- [51] J. A. Lenio, "Analysis of the Transtheoretical Model of behavior change," 2006.
- [52] J. O. Prochaska and C. C. DiClemente, "Transtheoretical therapy: Toward a more integrative model of change.," *Psychotherapy: theory, research & practice*, vol. 19, no. 3, p. 276, 1982.
- [53] M. Hashemzadeh, A. Rahimi, F. Zare-Farashbandi, A. M. Alavi-Naeini, and A. Daei, "Transtheoretical model of health behavioral change: A systematic review," *Iranian journal of nursing and midwifery research*, vol. 24, no. 2, p. 83, 2019.
- [54] H. Akbar, D. Anderson, and D. Gallegos, "Predicting intentions and behaviours in populations with or at-risk of diabetes: A systematic review," *Preventive medicine reports*, vol. 2, pp. 270–282, 2015.
- [55] I. Ajzen, "The theory of planned behavior: Frequently asked questions," *Human Behavior and Emerging Technologies*, vol. 2, no. 4, pp. 314–324, 2020.
- [56] I. Ajzen, "The theory of planned behaviour: Reactions and reflections," *Psychology & health*, vol. 26, no. 9. Taylor & Francis, pp. 1113–1127, 2011.
- [57] R. M. Ryan and E. L. Deci, "Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being.," *American psychologist*, vol. 55, no. 1, p. 68, 2000.
- [58] E. L. Deci and R. M. Ryan, "Self-determination theory: A macrotheory of human motivation, development, and health.," *Canadian psychology/Psychologie canadienne*, vol. 49, no. 3, p. 182, 2008.
- [59] J. De Man et al., "What motivates people with (pre) diabetes to move? Testing self-determination theory in rural Uganda," *Frontiers in Psychology*, vol. 11, p. 404, 2020.
- [60] A. S. Phillips and C. A. Guarnaccia, "Self-determination theory and motivational interviewing interventions for type 2 diabetes prevention and treatment: a systematic review," *Journal of health psychology*, vol. 25, no. 1, pp. 44–66, 2020.
- [61] R. Koestner and N. Hope, "A self-determination theory approach to goals," *The Oxford handbook of work engagement, motivation, and self-determination theory*, pp. 400–413, 2014.
- [62] K. Glanz, B. K. Rimer, and K. Viswanath, *Health behavior: Theory, research, and practice*. John Wiley & Sons, 2015.
- [63] E. C. Green, E. M. Murphy, and K. Gryboski, "The health belief model," *The Wiley encyclopedia of health psychology*, pp. 211–214, 2020.
- [64] M. Conner and P. Norman, "Social cognition models in health psychology." Taylor & Francis, 1998.
- [65] S. W. S. Lo, S. Y. Chair, and F. K. Lee, "Factors associated with health-promoting behavior of people with or at high risk of metabolic syndrome: based on the health belief model," *Applied Nursing Research*, vol. 28, no. 2, pp. 197–201, 2015.
- [66] A. Gopalan, I. S. Lorincz, C. Wirtalla, S. C. Marcus, and J. A. Long, "Awareness of prediabetes and engagement in diabetes risk-reducing behaviors," *American journal of preventive medicine*, vol. 49, no. 4, pp. 512–519, 2015.
- [67] R. G. Lord, J. M. Diefendorff, A. M. Schmidt, and R. J. Hall, "Self-regulation at work," *Annual review of psychology*, vol. 61, pp. 543–568, 2010.
- [68] S. Sultan, C. Attali, S. Gilberg, F. Zenasni, and A. Hartemann, "Physicians' understanding of patients' personal representations of their diabetes: accuracy and association with self-care," *Psychology & health*, vol. 26, no. suppl, pp. 101–117, 2011.
- [69] S. Sullivan-Bolyai et al., "Tried and true: self-regulation theory as a guiding framework for teaching parents diabetes education using human patient simulation," *ANS. Advances in nursing science*, vol. 37, no. 4, p. 340, 2014.
- [70] J. Troughton, J. Jarvis, C. Skinner, N. Robertson, K. Khunti, and M. Davies, "Waiting for diabetes: perceptions of people with pre-diabetes: a qualitative study," *Patient education and counseling*, vol. 72, no. 1, pp. 88–93, 2008.
- [71] K. W. Watkins, C. M. Connell, J. T. Fitzgerald, L. Klem, T. Hickey, and B. Ingersoll-Dayton, "Effect of adults' self-regulation of diabetes on quality-of-life outcomes.," *Diabetes care*, vol. 23, no. 10, pp. 1511–1515, 2000.
- [72] A. P. Cotter, N. Durant, A. A. Agne, and A. L. Cherrington, "Internet interventions to support lifestyle modification for diabetes management: a systematic review of the evidence," *Journal of Diabetes and its Complications*, vol. 28, no. 2, pp. 243–251, 2014.
- [73] E. J. Lyons, Z. H. Lewis, B. G. Mayrsohn, and J. L. Rowland, "Behavior change techniques implemented in electronic lifestyle activity monitors: a systematic content analysis," *Journal of medical Internet research*, vol. 16, no. 8, p. e192, 2014.
- [74] A. Bandura, "Health promotion from the perspective of social cognitive theory," in *Understanding and changing health behaviour*, Psychology Press, 2013, pp. 299–339.
- [75] B. H. Marcus and L. H. Forsyth, *Motivating people to be physically active*. Human Kinetics, 2008.
- [76] D. O'Sullivan and D. R. Strauser, "Operationalizing self-efficacy, related social cognitive variables, and moderating effects: implications for rehabilitation research and practice," *Rehabilitation Counseling Bulletin*, vol. 52, no. 4, pp. 251–258, 2009.
- [77] E. S. Anderson-Bill, R. A. Winnett, and J. R. Wojcik, "Social cognitive determinants of nutrition and physical activity among web-health users enrolling in an online intervention: the influence of social support, self-efficacy, outcome expectations, and self-regulation," *Journal of medical Internet research*, vol. 13, no. 1, p. e1551, 2011.
- [78] M. Klein, N. Mogles, and A. van Wissen, "An intelligent coaching system for therapy adherence," *IEEE pervasive computing*, vol. 12, no. 3, pp. 22–30, 2013.

- [79] J. D. Fisher and W. A. Fisher, "Changing AIDS-risk behavior.," *Psychological bulletin*, vol. 111, no. 3, p. 455, 1992.
- [80] W. A. Fisher, J. D. Fisher, and J. Harman, "The information-motivation-behavioral skills model: A general social psychological approach to understanding and promoting health behavior.," *Social psychological foundations of health and illness*, pp. 82–106, 2003.
- [81] L. S. Mayberry, R. L. Rothman, and C. Y. Osborn, "Family members' obstructive behaviors appear to be more harmful among adults with type 2 diabetes and limited health literacy.," *Journal of health communication*, vol. 19, no. sup2, pp. 132–143, 2014.
- [82] W. A. Fisher, J. D. Fisher, and J. Harman, "The information-motivation-behavioral skills model: A general social psychological approach to understanding and promoting health behavior.," *Social psychological foundations of health and illness*, pp. 82–106, 2003.
- [83] C. Y. Osborn, K. Rivet Amico, W. A. Fisher, L. E. Egede, and J. D. Fisher, "An information-motivation-behavioral skills analysis of diet and exercise behavior in Puerto Ricans with diabetes.," *Journal of health psychology*, vol. 15, no. 8, pp. 1201–1213, 2010.
- [84] J. Gao, J. Wang, Y. Zhu, and J. Yu, "Validation of an information-motivation-behavioral skills model of self-care among Chinese adults with type 2 diabetes.," *BMC Public Health*, vol. 13, no. 1, pp. 1–6, 2013.
- [85] E. Jeon and H.-A. Park, "Development of the IMB model and an evidence-based diabetes self-management mobile application.," *Healthcare informatics research*, vol. 24, no. 2, pp. 125–138, 2018.
- [86] T. Alanzi, R. Istepanian, and N. Philip, "Design and Usability Evaluation of Social Mobile Diabetes Management System in the Gulf Region.," *JMIR Res Protoc*, vol. 5, no. 3, p. e93, Sep. 2016, doi: 10.2196/resprot.4348.
- [87] M. M. Alotaibi, R. Istepanian, and N. Philip, "A mobile diabetes management and educational system for type-2 diabetics in Saudi Arabia (SAED).," *Mhealth*, vol. 2, 2016.
- [88] E. J. Aguiar, P. J. Morgan, C. E. Collins, R. C. Plotnikoff, M. D. Young, and R. Callister, "Efficacy of the type 2 diabetes prevention using lifestyle education program RCT.," *American journal of preventive medicine*, vol. 50, no. 3, pp. 353–364, 2016.
- [89] V. H. Buss, M. Varnfield, M. Harris, and M. Barr, "A mobile app for prevention of cardiovascular disease and type 2 diabetes mellitus: development and usability study.," *JMIR Human Factors*, vol. 9, no. 2, p. e35065, 2022.
- [90] M. D. Adu, U. H. Malabu, A. E. Malau-Aduli, and B. S. Malau-Aduli, "The development of My Care Hub mobile-phone app to support self-management in Australians with type 1 or type 2 diabetes.," *Scientific reports*, vol. 10, no. 1, p. 7, 2020.
- [91] G. Block et al., "Diabetes prevention and weight loss with a fully automated behavioral intervention by email, web, and mobile phone: a randomized controlled trial among persons with prediabetes.," *Journal of medical Internet research*, vol. 17, no. 10, p. e240, 2015.
- [92] C. M. Castro Sweet et al., "Outcomes of a digital health program with human coaching for diabetes risk reduction in a Medicare population.," *Journal of aging and health*, vol. 30, no. 5, pp. 692–710, 2018.
- [93] E. Cha et al., "A feasibility study to develop a diabetes prevention program for young adults with prediabetes by using digital platforms and a handheld device.," *The Diabetes Educator*, vol. 40, no. 5, pp. 626–637, 2014.
- [94] E. Everett, B. Kane, A. Yoo, A. Dobs, and N. Mathioudakis, "A novel approach for fully automated, personalized health coaching for adults with prediabetes: pilot clinical trial.," *Journal of medical Internet research*, vol. 20, no. 2, p. e72, 2018.
- [95] H. H. Fischer et al., "Text message support for weight loss in patients with prediabetes: a randomized clinical trial.," *Diabetes care*, vol. 39, no. 8, pp. 1364–1370, 2016.
- [96] Y. Fukuoka, C. L. Gay, K. L. Joiner, and E. Vittinghoff, "A novel diabetes prevention intervention using a mobile app: a randomized controlled trial with overweight adults at risk.," *American journal of preventive medicine*, vol. 49, no. 2, pp. 223–237, 2015.
- [97] M. K. Kramer et al., "A novel approach to diabetes prevention: evaluation of the Group Lifestyle Balance program delivered via DVD.," *Diabetes research and clinical practice*, vol. 90, no. 3, pp. e60–e63, 2010.
- [98] J. Ma et al., "Translating the Diabetes Prevention Program lifestyle intervention for weight loss into primary care: a randomized trial.," *JAMA internal medicine*, vol. 173, no. 2, pp. 113–121, 2013.
- [99] A. Michaelides, C. Raby, M. Wood, K. Farr, and T. Toro-Ramos, "Weight loss efficacy of a novel mobile Diabetes Prevention Program delivery platform with human coaching.," *BMJ Open Diabetes Research and Care*, vol. 4, no. 1, p. e000264, 2016.
- [100] G. A. Piatt, M. C. Seidel, R. O. Powell, and J. C. Zgibor, "Comparative effectiveness of lifestyle intervention efforts in the community: results of the Rethinking Eating and ACTivity (REACT) study.," *Diabetes Care*, vol. 36, no. 2, pp. 202–209, 2013.
- [101] S. C. Sepah, L. Jiang, and A. L. Peters, "Translating the diabetes prevention program into an online social network: validation against CDC standards.," *The Diabetes Educator*, vol. 40, no. 4, pp. 435–443, 2014.
- [102] M. G. Wilson et al., "Evaluation of a digital behavioral counseling program for reducing risk factors for chronic disease in a workforce.," *Journal of occupational and environmental medicine*, vol. 59, no. 8, p. e150, 2017.
- [103] J. H. Arens, W. Hauth, and J. Weissmann, "Novel app-and web-supported diabetes prevention program to promote weight reduction, physical activity, and a healthier lifestyle: observation of the clinical application.," *Journal of diabetes science and technology*, vol. 12, no. 4, pp. 831–838, 2018.
- [104] T. Limaye et al., "Efficacy of a virtual assistance-based lifestyle intervention in reducing risk factors for Type 2 diabetes in young employees in the information technology industry in India: LIMIT, a randomized controlled trial.," *Diabetic medicine*, vol. 34, no. 4, pp. 563–568, 2017.
- [105] A. Ramachandran et al., "Effectiveness of mobile phone messaging in prevention of type 2 diabetes by lifestyle modification in men in India: a prospective, parallel-group, randomised controlled trial.," *The lancet Diabetes & endocrinology*, vol. 1, no. 3, pp. 191–198, 2013.
- [106] A. Rose, B. M. Deros, and M. A. Rahman, "A study on lean manufacturing implementation in Malaysian automotive component industry.," *International Journal of Automotive and Mechanical Engineering*, vol. 8, pp. 1467–1476, 2013.
- [107] A. M. Boels, R. C. Vos, L.-T. Dijkhorst-Oei, and G. E. Rutten, "Effectiveness of diabetes self-management education and support via a smartphone application in insulin-treated patients with type 2 diabetes: results of a randomized controlled trial (TRIGGER study).," *BMJ Open Diabetes Research and Care*, vol. 7, no. 1, p. e000981, 2019.
- [108] K. A. Cradock et al., "Design of a planner-based intervention to facilitate diet behaviour change in type 2 diabetes.," *Sensors*, vol. 22, no. 7, p. 2795, 2022.
- [109] D. Y. Chao, T. M. Lin, and W.-Y. Ma, "Enhanced self-efficacy and behavioral changes among patients with diabetes: cloud-based mobile health platform and mobile app service.," *JMIR diabetes*, vol. 4, no. 2, p. e11017, 2019.
- [110] C. Sun et al., "Mobile phone-based telemedicine practice in older chinese patients with type 2 diabetes mellitus: randomized controlled trial.," *JMIR mHealth and uHealth*, vol. 7, no. 1, p. e10664, 2019.
- [111] T. Alanzi, R. Istepanian, N. Philip, and others, "Design and usability evaluation of social mobile diabetes management system in the Gulf Region.," *JMIR research protocols*, vol. 5, no. 3, p. e4348, 2016.
- [112] E. Mehraeen et al., "Design and development of a mobile-based self-care application for patients with type 2 diabetes.," *Journal of Diabetes Science and Technology*, vol. 16, no. 4, pp. 1008–1015, 2022.
- [113] S. Subramaniam, J. S. Dhillon, and W. F. Wan Ahmad, "Behavioral Theory-Based Framework for Prediabetes Self-Care System—Design Perspectives and Validation Results.," *International journal of environmental research and public health*, vol. 18, no. 17, p. 9160, 2021.
- [114] D. H. Frøisland, E. Årsand, and F. Skårderud, "Improving diabetes care for young people with type 1 diabetes through visual learning on mobile phones: mixed-methods study.," *Journal of medical Internet research*, vol. 14, no. 4, p. e2155, 2012.
- [115] H. Holmen et al., "A mobile health intervention for self-management and lifestyle change for persons with type 2 diabetes, part 2: one-year

- results from the Norwegian randomized controlled trial RENEWING HEALTH,” JMIR mHealth and uHealth, vol. 2, no. 4, p. e3882, 2014.
- [116] R. A. Sowah, A. A. Bampoe-Addo, S. K. Armoo, F. K. Saalia, F. Gatsi, and B. Sarkodie-Mensah, “Design and development of diabetes management system using machine learning,” International journal of telemedicine and applications, vol. 2020, 2020.
- [117] Kusananto, K. A. J. Widyana, Suprajitno, and H. Arifin, “DM-calendar app as a diabetes self-management education on adult type 2 diabetes mellitus: a randomized controlled trial,” Journal of Diabetes & Metabolic Disorders, vol. 18, pp. 557–563, 2019.
- [118] L. Ledderer, A. Møller, and A. Fage-Butler, “Adolescents’ participation in their healthcare: A sociomaterial investigation of a diabetes app,” Digital health, vol. 5, p. 2055207619845448, 2019.
- [119] K. Waki et al., “DialBetics: a novel smartphone-based self-management support system for type 2 diabetes patients,” Journal of diabetes science and technology, vol. 8, no. 2, pp. 209–215, 2014.
- [120] C. C. Quinn, M. D. Shardell, M. L. Terrin, E. A. Barr, S. H. Ballew, and A. L. Gruber-Baldini, “Cluster-randomized trial of a mobile phone personalized behavioral intervention for blood glucose control,” Diabetes care, vol. 34, no. 9, pp. 1934–1942, 2011.
- [121] K. A. Cradock, G. ÓLaighin, F. M. Finucane, H. L. Gainforth, L. R. Quinlan, and K. A. M. Ginis, “Behaviour change techniques targeting both diet and physical activity in type 2 diabetes: A systematic review and meta-analysis,” International Journal of Behavioral Nutrition and Physical Activity, vol. 14, no. 1, pp. 1–17, 2017.
- [122] H. Fu, S. K. McMahon, C. R. Gross, T. J. Adam, and J. F. Wyman, “Usability and clinical efficacy of diabetes mobile applications for adults with type 2 diabetes: A systematic review,” Diabetes research and clinical practice, vol. 131, pp. 70–81, 2017.
- [123] J. Singh, B. C. Wünsche, and C. Lutteroth, “Framework for Healthcare4Life: a ubiquitous patient-centric telehealth system,” in Proceedings of the 11th International Conference of the NZ Chapter of the ACM Special Interest Group on Human-Computer Interaction, 2010, pp. 41–48.
- [124] S. Subramaniam, J. S. Dhillon, and W. F. Wan Ahmad, “Behavioral Theory-Based Framework for Prediabetes Self-Care System—Design Perspectives and Validation Results,” IJERPH, vol. 18, no. 17, p. 9160, Aug. 2021, doi: 10.3390/ijerph18179160.
- [125] G. Block et al., “Diabetes prevention and weight loss with a fully automated behavioral intervention by email, web, and mobile phone: a randomized controlled trial among persons with prediabetes,” Journal of medical Internet research, vol. 17, no. 10, p. e240, 2015.
- [126] A. Kankanhalli, J. Shin, H. Oh, and others, “Mobile-based interventions for dietary behavior change and health outcomes: scoping review,” JMIR mHealth and uHealth, vol. 7, no. 1, p. e11312, 2019.
- [127] E. Jeon and H.-A. Park, “Experiences of Patients With a Diabetes Self-Care App Developed Based on the Information-Motivation-Behavioral Skills Model: Before-and-After Study,” JMIR Diabetes, vol. 4, no. 2, p. e11590, Apr. 2019, doi: 10.2196/11590.
- [128] Y. Wu et al., “Mobile App-Based Interventions to Support Diabetes Self-Management: A Systematic Review of Randomized Controlled Trials to Identify Functions Associated with Glycemic Efficacy,” JMIR Mhealth Uhealth, vol. 5, no. 3, p. e35, Mar. 2017, doi: 10.2196/mhealth.6522.
- [129] B. C. Bonoto et al., “Efficacy of mobile apps to support the care of patients with diabetes mellitus: a systematic review and meta-analysis of randomized controlled trials,” JMIR mHealth and uHealth, vol. 5, no. 3, p. e6309, 2017.
- [130] S. Michie et al., “The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: Building an international consensus for the reporting of behavior change interventions,” Annals of Behavioral Medicine, vol. 46, no. 1, pp. 81–95, Aug. 2013, doi: 10.1007/s12160-013-9486-6.
- [131] S. Michie, M. M. Van Stralen, and R. West, “The behaviour change wheel: A new method for characterising and designing behaviour change interventions,” Implementation Sci, vol. 6, no. 1, p. 42, Dec. 2011, doi: 10.1186/1748-5908-6-42.
- [132] K. L. Joiner, S. Nam, and R. Whittemore, “Lifestyle interventions based on the diabetes prevention program delivered via eHealth: a systematic review and meta-analysis,” Preventive medicine, vol. 100, pp. 194–207, 2017.
- [133] R. R. Bian et al., “The effect of technology-mediated diabetes prevention interventions on weight: a meta-analysis,” Journal of medical Internet research, vol. 19, no. 3, p. e76, 2017.
- [134] V. Braun and V. Clarke, “Using thematic analysis in psychology,” Qualitative research in psychology, vol. 3, no. 2, pp. 77–101, 2006.
- [135] P. Ince, G. Haddock, and S. Tai, “A systematic review of the implementation of recommended psychological interventions for schizophrenia: rates, barriers, and improvement strategies,” Psychology and Psychotherapy: Theory, Research and Practice, vol. 89, no. 3, pp. 324–350, 2016.
- [136] M. Taloyan, M. Kia, F. Lamian, M. Peterson, and E. Rydwik, “Web-based support for individuals with type 2 diabetes—a feasibility study,” BMC Health Services Research, vol. 21, no. 1, pp. 1–8, 2021.
- [137] Y. Wang et al., “Effects of continuous care for patients with type 2 diabetes using mobile health application: a randomised controlled trial,” The International journal of health planning and management, vol. 34, no. 3, pp. 1025–1035, 2019.
- [138] B. E. Holtz et al., “The design and development of MyTIDHero: A mobile app for adolescents with type 1 diabetes and their parents,” J Telemed Telecare, vol. 25, no. 3, pp. 172–180, Apr. 2019, doi: 10.1177/1357633X17745470.
- [139] M. Afarideh et al., “Complex association of serum alanine aminotransferase with the risk of future cardiovascular disease in type 2 diabetes,” Atherosclerosis, vol. 254, pp. 42–51, 2016.
- [140] E. Cha et al., “Health literacy, self-efficacy, food label use, and diet in young adults,” American journal of health behavior, vol. 38, no. 3, pp. 331–339, 2014.
- [141] E. J. Aguiar, P. J. Morgan, C. E. Collins, R. C. Plotnikoff, M. D. Young, and R. Callister, “Efficacy of the Type 2 Diabetes Prevention Using LifeStyle Education Program RCT,” American Journal of Preventive Medicine, vol. 50, no. 3, pp. 353–364, 2016, doi: 10.1016/j.amepre.2015.08.020.
- [142] M. D. Adu, U. H. Malabu, A. E. O. Malau-Aduli, and B. S. Malau-Aduli, “The development of My Care Hub Mobile-Phone App to Support Self-Management in Australians with Type 1 or Type 2 Diabetes,” Scientific Reports, vol. 10, no. 1, pp. 1–10, 2020, doi: 10.1038/s41598-019-56411-0.
- [143] G. Block et al., “Diabetes prevention and weight loss with a fully automated behavioral intervention by email, web, and mobile phone: A randomized controlled trial among persons with prediabetes,” Journal of Medical Internet Research, vol. 17, no. 10, p. e4897, 2015, doi: 10.2196/jmir.4897.
- [144] S. Subramaniam, J. S. Dhillon, and W. F. Wan Ahmad, “Behavioral theory-based framework for prediabetes self-care system—design perspectives and validation results,” International Journal of Environmental Research and Public Health, vol. 18, no. 17, 2021, doi: 10.3390/ijerph18179160.
- [145] C. M. Castro Sweet et al., “Outcomes of a Digital Health Program With Human Coaching for Diabetes Risk Reduction in a Medicare Population,” Journal of Aging and Health, vol. 30, no. 5, pp. 692–710, 2018, doi: 10.1177/0898264316688791.
- [146] A. M. Boels, R. C. Vos, L. T. Dijkhorst-Oei, and G. E. H. M. Rutten, “Effectiveness of diabetes self-management education and support via a smartphone application in insulin-treated patients with type 2 diabetes: Results of a randomized controlled trial (TRIGGER study),” BMJ Open Diabetes Research and Care, vol. 7, no. 1, pp. 1–4, 2019, doi: 10.1136/bmjdr-2019-000981.
- [147] M. Afarideh, A. Ghajar, S. Noshad, and A. Esteghamati, “Text message support for weight loss in patients with prediabetes: A randomized clinical trial,” Diabetes Care, vol. 39, no. 11, p. e206, 2016, doi: 10.2337/dc16-1210.
- [148] B. E. Holtz et al., “The design and development of MyTIDHero: A mobile app for adolescents with type 1 diabetes and their parents,” Journal of Telemedicine and Telecare, vol. 25, no. 3, pp. 172–180, Dec. 2019, doi: 10.1177/1357633X17745470.
- [149] A. M. Boels, R. C. Vos, L. T. Dijkhorst-Oei, and G. E. H. M. Rutten, “Effectiveness of diabetes self-management education and support via a

- smartphone application in insulin-treated patients with type 2 diabetes: Results of a randomized controlled trial (TRIGGER study),” *BMJ Open Diabetes Research and Care*, vol. 7, no. 1, pp. 1–4, 2019, doi: 10.1136/bmjdr-2019-000981.
- [150] K. A. Cradock et al., “Design of a Planner-Based Intervention to Facilitate Diet Behaviour Change in Type 2 Diabetes,” *Sensors*, vol. 22, no. 7, pp. 1–28, 2022, doi: 10.3390/s22072795.
- [151] E. Everett, B. Kane, A. Yoo, A. Dobs, and N. Mathioudakis, “A Novel Approach for Fully Automated, Personalized Health Coaching for Adults with Prediabetes: Pilot Clinical Trial,” *Journal of medical Internet research*, vol. 20, no. 2, p. e72, 2018, doi: 10.2196/jmir.9723.
- [152] Kusnanto, K. A. J. Widyana, Suprajitno, and H. Arifin, “DM-calendar app as a diabetes self-management education on adult type 2 diabetes mellitus: a randomized controlled trial,” *Journal of Diabetes and Metabolic Disorders*, vol. 18, no. 2, pp. 557–563, 2019, doi: 10.1007/s40200-019-00468-1.
- [153] C. M. Castro Sweet et al., “Outcomes of a Digital Health Program With Human Coaching for Diabetes Risk Reduction in a Medicare Population,” *Journal of Aging and Health*, vol. 30, no. 5, pp. 692–710, 2018, doi: 10.1177/0898264316688791.
- [154] E. Cha et al., “A Feasibility Study to Develop a Diabetes Prevention Program for Young Adults With Prediabetes by Using Digital Platforms and a Handheld Device,” *The Diabetes Educator*, vol. 40, no. 5, pp. 626–637, 2014, doi: 10.1177/0145721714539736.
- [155] H. H. Fischer et al., “Text message support for weight loss in patients with prediabetes: A randomized clinical trial,” *Diabetes Care*, vol. 39, no. 11, p. e206, 2016, doi: 10.2337/dc16-1210.
- [156] Y. Fukuoka, C. L. Gay, K. L. Joiner, and E. Vittinghoff, “A Novel Diabetes Prevention Intervention Using a Mobile App,” *American Journal of Preventive Medicine*, vol. 49, no. 2, pp. 223–237, 2015, doi: 10.1016/j.amepre.2015.01.003.
- [157] M. K. Kramer et al., “A novel approach to diabetes prevention: Evaluation of the Group Lifestyle Balance program delivered via DVD,” *Diabetes Research and Clinical Practice*, vol. 90, no. 3, pp. e60–e63, 2010, doi: 10.1016/j.diabres.2010.08.013.
- [158] A. Michaelides, C. Raby, M. Wood, K. Farr, and T. Toro-Ramos, “Weight loss efficacy of a novel mobile diabetes prevention program delivery platform with human coaching,” *BMJ Open Diabetes Research and Care*, vol. 4, no. 1, p. e000264, 2016, doi: 10.1136/bmjdr-2016-000264.
- [159] R. R. Bian et al., “The effect of technology-mediated diabetes prevention interventions on weight: A meta-analysis,” *Journal of Medical Internet Research*, vol. 19, no. 3, p. e4709, 2017, doi: 10.2196/jmir.4709.
- [160] G. A. Piatt, M. C. Seidel, R. O. Powell, and J. C. Zgibor, “Comparative effectiveness of lifestyle intervention efforts in the community: Results of the rethinking eating and ACTivity (REACT) study,” *Diabetes Care*, vol. 36, no. 2, pp. 202–209, 2013, doi: 10.2337/dc12-0824.
- [161] S. C. Sepah, L. Jiang, and A. L. Peters, “Translating the Diabetes Prevention Program into an Online Social Network: Validation against CDC Standards,” *The Diabetes Educator*, vol. 40, no. 4, pp. 435–443, 2014, doi: 10.1177/0145721714531339.
- [162] M. G. Wilson et al., “Evaluation of a Digital Behavioral Counseling Program for Reducing Risk Factors for Chronic Disease in a Workforce,” *Journal of Occupational and Environmental Medicine*, vol. 59, no. 8, pp. e150–e155, 2017, doi: 10.1097/JOM.0000000000001091.
- [163] E. Mehraeen et al., “Design and Development of a Mobile-Based Self-Care Application for Patients with Type 2 Diabetes,” *Journal of Diabetes Science and Technology*, vol. 16, no. 4, pp. 1008–1015, 2022, doi: 10.1177/19322968211007124.
- [164] C. K. H. Wong et al., “A short message service (SMS) intervention to prevent diabetes in Chinese professional drivers with pre-diabetes: A pilot single-blinded randomized controlled trial,” *Diabetes Research and Clinical Practice*, vol. 102, no. 3, pp. 158–166, Dec. 2013, doi: 10.1016/j.diabres.2013.10.002.
- [165] A. L. Orsama et al., “Active assistance technology reduces glycosylated hemoglobin and weight in individuals with type 2 diabetes: Results of a theory-based randomized trial,” *Diabetes Technology and Therapeutics*, vol. 16, no. SUPPL. 1, 2014, doi: 10.1089/dia.2014.1507.
- [166] J. Ma et al., “Translating the diabetes prevention program lifestyle intervention for weight loss into primary care: A randomized trial,” *JAMA Internal Medicine*, vol. 173, no. 2, pp. 113–121, 2013, doi: 10.1001/2013.jamainternmed.987.
- [167] A. Ramachandran et al., “Effectiveness of mobile phone messaging in prevention of type 2 diabetes by lifestyle modification in men in India: A prospective, parallel-group, randomized controlled trial,” *Diabetes Technology and Therapeutics*, vol. 17, no. 3, pp. S65–S66, 2015, doi: 10.1089/dia.2015.1507.
- [168] T. Alanzi, R. Istepanian, and N. Philip, “Design and Usability Evaluation of Social Mobile Diabetes Management System in the Gulf Region,” *JMIR Research Protocols*, vol. 5, no. 3, p. e93, 2016, doi: 10.2196/resprot.4348.
- [169] H. Holmen et al., “A mobile health intervention for self-management and lifestyle change for persons with type 2 diabetes, part 2: One-year results from the norwegian randomized controlled trial RENEWING HEALTH,” *JMIR mHealth and uHealth*, vol. 2, no. 4, pp. 1–16, 2014, doi: 10.2196/mhealth.3882.
- [170] L. Ledderer, A. Møller, and A. Fage-Butler, “Adolescents’ participation in their healthcare: A sociomaterial investigation of a diabetes app,” *Digital Health*, vol. 5, Apr. 2019, doi: 10.1177/2055207619845448.
- [171] V. H. Buss, M. Varnfield, M. Harris, and M. Barr, “A Mobile App for Prevention of Cardiovascular Disease and Type 2 Diabetes Mellitus: Development and Usability Study,” *JMIR Human Factors*, vol. 9, no. 2, pp. 1–20, 2022, doi: 10.2196/35065.
- [172] G. Block et al., “Diabetes prevention and weight loss with a fully automated behavioral intervention by email, web, and mobile phone: A randomized controlled trial among persons with prediabetes,” *Journal of Medical Internet Research*, vol. 17, no. 10, p. e4897, 2015, doi: 10.2196/jmir.4897.
- [173] A. Michaelides, C. Raby, M. Wood, K. Farr, and T. Toro-Ramos, “Weight loss efficacy of a novel mobile diabetes prevention program delivery platform with human coaching,” *BMJ Open Diabetes Research and Care*, vol. 4, no. 1, p. e000264, 2016, doi: 10.1136/bmjdr-2016-000264.
- [174] S. C. Sepah, L. Jiang, and A. L. Peters, “Translating the Diabetes Prevention Program into an Online Social Network: Validation against CDC Standards,” *The Diabetes Educator*, vol. 40, no. 4, pp. 435–443, 2014, doi: 10.1177/0145721714531339.
- [175] D. Y. P. Chao, T. M. Y. Lin, and W. Y. Ma, “Enhanced self-efficacy and behavioral changes among patients with diabetes: Cloud-based mobile health platform and mobile app service,” *JMIR Diabetes*, vol. 4, no. 2, 2019, doi: 10.2196/11017.
- [176] T. L. Webb, J. Joseph, L. Yardley, and S. Michie, “Using the Internet to promote health behavior change: A systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy,” *Journal of Medical Internet Research*, vol. 12, no. 1, p. e1376, Feb. 2010, doi: 10.2196/jmir.1376.
- [177] L. Van Rhoon, M. Byrne, E. Morrissey, J. Murphy, and J. McSharry, “A systematic review of the behaviour change techniques and digital features in technology-driven type 2 diabetes prevention interventions,” *Digital Health*, vol. 6, 2020, doi: 10.1177/2055207620914427.
- [178] D. Y. P. Chao, T. M. Y. Lin, and W. Y. Ma, “Enhanced self-efficacy and behavioral changes among patients with diabetes: Cloud-based mobile health platform and mobile app service,” *JMIR Diabetes*, vol. 4, no. 2, 2019, doi: 10.2196/11017.
- [179] Kusnanto, K. A. J. Widyana, Suprajitno, and H. Arifin, “DM-calendar app as a diabetes self-management education on adult type 2 diabetes mellitus: a randomized controlled trial,” *Journal of Diabetes and Metabolic Disorders*, vol. 18, no. 2, pp. 557–563, 2019, doi: 10.1007/s40200-019-00468-1.
- [180] P. Craig, P. Dieppe, S. Macintyre, S. Michie, I. Nazareth, and M. Petticrew, “Developing and evaluating complex interventions: the new Medical Research Council guidance,” *Bmj*, vol. 337, 2008.
- [181] S. Michie and A. Prestwich, “Are interventions theory-based? Development of a theory coding scheme,” *Health psychology*, vol. 29, no. 1, p. 1, 2010.



- [182] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, 2006, doi: 10.1191/1478088706qp0630a.
- [183] S. B. Donevant, R. D. Estrada, J. M. Culley, B. Habing, and S. A. Adams, "Exploring app features with outcomes in mHealth studies involving chronic respiratory diseases, diabetes, and hypertension: a targeted exploration of the literature," *Journal of the American Medical Informatics Association*, vol. 25, no. 10, pp. 1407–1418, Oct. 2018, doi: 10.1093/jamia/ocy104.
- [184] L. S. Holcomb, "A taxonomic integrative review of short message service (SMS) methodology: a framework for improved diabetic outcomes," *Journal of diabetes science and technology*, vol. 9, no. 6, pp. 1321–1326, 2015.
- [185] J. Lindström et al., "Take action to prevent diabetes- The IMAGE toolkit for the prevention of type 2 diabetes in Europe," *Hormone and Metabolic Research*, vol. 42, no. SUPPL. 1, pp. S37–S55, 2010, doi: 10.1055/s-0029-1240975.
- [186] S. Michie et al., "The Behavior Change Technique Taxonomy ( v1 ) of 93 Hierarchically Clustered Techniques : Building an International Consensus for the Reporting of Behavior Change Interventions," pp. 81–95, 2013, doi: 10.1007/s12160-013-9486-6.
- [187] Q. Yang and S. K. Van Stee, "The comparative effectiveness of mobile phone interventions in improving health outcomes: Meta-analytic review," *JMIR mHealth and uHealth*, vol. 7, no. 4, pp. 1–14, 2019, doi: 10.2196/11244.
- [188] K. Liu, Z. Xie, and C. K. Or, "Effectiveness of mobile app-assisted self-care interventions for improving patient outcomes in type 2 diabetes and/or hypertension: Systematic review and meta-analysis of randomized controlled trials," *JMIR mHealth and uHealth*, vol. 8, no. 8, pp. 1–23, 2020, doi: 10.2196/15779.
- [189] Y. Mao, W. Lin, J. Wen, and G. Chen, "Impact and efficacy of mobile health intervention in the management of diabetes and hypertension : a systematic analysis review and meta," pp. 1–11, 2020, doi: 10.1136/bmjdr-2020-001225.
- [190] Y. Wu et al., "Mobile app-based interventions to support diabetes self-management: A systematic review of randomized controlled trials to identify functions associated with glycemic efficacy," *JMIR mHealth and uHealth*, vol. 5, no. 3, 2017, doi: 10.2196/mhealth.6522.
- [191] X. Wu, X. Guo, and Z. Zhang, "The efficacy of mobile phone apps for lifestyle modification in diabetes: Systematic review and meta-analysis," *JMIR mHealth and uHealth*, vol. 7, no. 1, pp. 1–13, 2019, doi: 10.2196/12297.
- [192] J. P. Riley, J. P. Gabe, and M. R. Cowie, "Does telemonitoring in heart failure empower patients for self-care? A qualitative study," *J Clin Nurs*, vol. 22, no. 17–18, pp. 2444–2455, Sep. 2013, doi: 10.1111/j.1365-2702.2012.04294.x.
- [193] R. R. Bian et al., "The effect of technology-mediated diabetes prevention interventions on weight: A meta-analysis," *Journal of Medical Internet Research*, vol. 19, no. 3, p. e4709, 2017, doi: 10.2196/jmir.4709.
- [194] Y. Mao, W. Lin, J. Wen, and G. Chen, "Impact and efficacy of mobile health intervention in the management of diabetes and hypertension : a systematic analysis review and meta," pp. 1–11, 2020, doi: 10.1136/bmjdr-2020-001225.
- [195] A. J. Dunkley et al., "Diabetes prevention in the real world: Effectiveness of pragmatic lifestyle interventions for the prevention of type 2 diabetes and of the impact of adherence to guideline recommendations: A systematic review and meta-analysis (Diabetes Care 201)," *Diabetes Care*, vol. 37, no. 6, pp. 1775–1776, 2014, doi: 10.2337/dc14-er06.
- [196] J. E. Donnelly, S. N. Blair, J. M. Jakicic, M. M. Manore, J. W. Rankin, and B. K. Smith, "Appropriate physical activity intervention strategies for weight loss and prevention of weight regain for adults," *Medicine and Science in Sports and Exercise*, vol. 41, no. 2, pp. 459–471, 2009, doi: 10.1249/MSS.0b013e3181949333.
- [197] J. A. Dunbar et al., "Public Health Approaches to Type 2 Diabetes Prevention: The US National Diabetes Prevention Program and Beyond," *Current Diabetes Reports*, vol. 19, no. 11, pp. 1–11, 2019, doi: 10.1007/s11892-019-1262-y.

# An Improved Convolutional Neural Network for Churn Analysis

Priya Gopal, Dr. Nazri Bin MohdNawi

Faculty of Computer Science and Information Technology, University Tun Hussein Onn Malaysia

**Abstract**—The significance of customer churn analysis has escalated due to the increasing availability of relevant data and intensifying competition. Researchers and practitioners are focused on enhancing prediction accuracy in modeling approaches, with deep neural networks emerging as appealing due to their robust performance across domains. However, the computational demands surge due to the challenges posed by dimensionality and inherent characteristics of the data. To address these issues, this research proposes a novel hybrid model that strategically integrates Convolutional Neural Networks (CNN) and a modified Variational Autoencoder (VAE). By carefully adjusting the parameters of the VAE to capture the central tendency and range of variation, the study aims to enhance the effectiveness of classifying high-dimensional churn data. The proposed framework's efficacy is evaluated using six benchmark datasets from various domains, with performance metrics encompassing accuracy, f1-score, precision, recall, and response time. Experimental results underscore the prowess of the hybrid technique in effectively handling high-dimensional and imbalanced time series data, thus offering a robust pathway for enhanced churn analysis.

**Keywords**—Customer churn analysis; deep learning; variational autoencoder; convolutional neural networks; dimensionality reduction

## I. INTRODUCTION

In today's rapidly evolving business landscape, driven by the surge of online technological advancements, companies are compelled to navigate a competitive arena characterized by the influx of new business models and market entrants [1]. This has intensified the significance of customer churn analysis, as businesses seek to attract new customers and retain their existing clientele [2]. Retaining customers has been proven to yield higher returns on investment, as the costs associated with retaining an existing customer are considerably lower than acquiring a new one [3]. Amidst this context, the retention strategy gains paramount importance, requiring companies to mitigate the risk of customer churn – the phenomenon where customers switch providers swiftly [4-5].

To address this challenge, the utilization of machine learning has emerged as a potent tool, leveraging historical data to predict potential churn events and enable informed decision-making [6, 7]. However, there are hurdles to overcome in this endeavor. Issues such as inaccurate customer information, intricate datasets with numerous variables, imbalanced class distributions, and a lack of industry expertise create formidable hurdles [8]. Despite the strides made by advanced techniques like Convolutional Neural Networks (CNNs), which uncover hidden relationships within data, accurately predicting real-

world churn scenarios remains intricate [9-11]. In light of these challenges, this paper introduces a hybrid model named the Space Vector Variational Autoencoder (SV-VAE), a fusion of CNN, and an optimized Variational Autoencoder (VAE) [12, 13].

By addressing these challenges, this study contributes to the enhancement of churn prediction in the dynamic landscape of modern business. It brings together cutting-edge technologies in a concerted effort to improve retention rates and elevate the strategic decision-making process for businesses across diverse industries.

The core objective of this paper is to enhance both the accuracy of predictions and the efficiency of model learning. This enhancement is achieved through the integration of a Convolutional Neural Network (CNN) with a modified Variational Autoencoder (SV-VAE). By combining these techniques, we aim to achieve superior performance in terms of predictive precision and reduced model training time.

To validate the effectiveness of the proposed hybrid model, a comprehensive evaluation is conducted. This evaluation encompasses various critical performance metrics, including precision, recall, accuracy, and learning time. To establish a robust baseline for comparison, the proposed SV-VAE hybrid model is benchmarked against other popular autoencoder architectures such as Vanilla, Stacked, Sparse, Denoising, and Variational Autoencoders. These comparisons are conducted across diverse industry-standard benchmark datasets, which provide a real-world context for assessing model performance.

The validation process primarily centers around the scrutiny of the proposed model's predictive capabilities. The study meticulously assesses the accuracy of predictions, the ability to accurately classify positive instances (precision), and the model's effectiveness in capturing actual positive instances (recall). This thorough evaluation ensures that the proposed hybrid SV-VAE model's performance improvements are statistically significant and practically relevant in the context of churn analysis and prediction tasks. The following section contains a comprehensive review of the existing literature in the field of churn prediction, machine learning techniques, and autoencoder architectures relevant to this study.

## II. LITERATURE REVIEW

Many methods have been explored in the quest to predict churn in service industries, often rooted in machine learning and data mining techniques. A significant portion of the prior research has been concentrated on individual data mining

techniques or has involved comparative analyses of different methodologies for predicting attrition.

In a study conducted by Brandusoiu et al. [11], the focus was on predicting prepaid customer turnover rates using a contemporary data mining approach. The study utilized a dataset comprising more than 3000 call details, encompassing 21 attributes, and a predictive churn variable categorized with Yes/No labels. These attributes encompassed details about the volume of voice and video usage for each subscriber, alongside the count of inbound and outbound texts. The researcher employed the principal component analysis algorithm to streamline the data's complexity for dimensionality reduction. Three machine learning algorithms, namely Support Vector Machine (SVM), Naive Bayes (NB), and Neural Networks (NN), were employed to forecast churn rates. Model reliability was assessed using the Area under the Curve metric, and the results highlighted the superior performance of SVM over the other two algorithms. Notably, the dataset used in this study didn't contain any missing values. However, when dealing with time-series features, the model's ability to leverage information over time might be limited, necessitating various sampling techniques to incorporate temporal information effectively [12-13].

Artificial neural network approaches designed for sequential data have gained popularity, and this trend is evident in their increased adoption for churn modeling, as evident from the overview provided in Table I.

TABLE I. CHURN MODELS ANALYZED BY DIFFERENT AUTHORS

Paper	NN technique	Industry Data	Accuracy
Nasebah et al. (2019) [17]	CNN, Modified – CNN	Telecom	Accuracy, precision, recall, F-measure, ROC & AUC
A. S. Kumara and D. Chandrakala (2016) [18]	LSTM, RFM + LSTM	Telecom	Mean evaluation metric
Domingos et al. (2021) [19]	MLP, DNN	Banking Sector	Accuracy using RMSProp, SGD, Adam
Ahmed et al. (2019) [20]	CNN Classifiers, custom CNN	Telecom	Prediction Accuracy, ROC
Zhou et al. (2019) [21]	DL-CNN, One-dimensional CNN, XGBoost	Online New Media Platform	Precision Recall
Umayaparvathi and Iyakutti et al. (2017) [22]	CNN, mall FNN, Large FNN	Telecom	Accuracy
Wangperawong et al. (2016)	Deep CNN, Autoencoder	Time-series data	AUC
Kristensen et al (2019)	LSTM, Aggregated LSTM, LSTM Hidden State	freemium games	ROC, AUC & Accuracy
Prosvetov and Artem. (2018) [25]	CNN-based autoencoder, LSTM-based autoencoder	Telecom	roc-auc metric

Martins [14] conducted a study revealing that the accuracy of Long Short-Term Memory (LSTM) models equipped with time series attributes is comparable to an approach that integrates this pertinent data using the mean and a random forest technique. This research outcome contributes to synthesizing insights from various studies.

Numerous research endeavors have underscored the efficacy of combining Convolutional Neural Networks (CNN) with Long Short-Term Memory (LSTM) models to enhance performance across diverse tasks. For example, in a study by Tan et al., [15], a fusion of CNN and LSTM was employed to forecast user-intended actions, demonstrating that this integration surpasses the individual performance of the two models and traditional machine learning methods. Furthermore, another investigation employed encoded sequential data, such as images and videos, wherein the utilization of CNN yielded superior outcomes compared to gradient boosting and random forest techniques [16].

These discoveries accentuate the potential advantages of harnessing the strengths of distinct models, leading to heightened performance across a spectrum of applications.

In summary, previous studies indicate that the exploration of time-varying data to enhance the effectiveness of churn algorithms is still in its nascent stages. While CNN and LSTM models exhibit improved performance, they encounter challenges when confronted with high-dimensional data. Moreover, the diverse characteristics inherent to various industry sectors make it challenging to definitively determine the performance boost resulting from the inclusion of such varied information. Introducing these diverse features into the training phase can inadvertently amplify model complexity, potentially leading to overfitting against the training data. A potential solution to this lies in the preprocessing stage, where a dimensionality reduction step is undertaken. This step strives to curtail the number of features while retaining as much meaningful information as possible within the dataset [25]. Autoencoders prove adept at handling high-dimensional data, a domain where CNNs might face limitations. Notably, the Variational Autoencoder (VAE) is of special significance, given its ability to generate more probabilistic latent outputs [18]. The VAE emerges as a robust choice, particularly well-suited for churn analysis due to its capability to generate novel data instances and its compatibility with the time series nature of churn data. Additionally, dimensionality reduction plays a vital role in effectively mitigating noise from the data. This process facilitates the discovery of latent variables that arise from intricate relationships among different variables in the dataset. This approach provides a more comprehensive understanding that goes beyond analyzing individual variables in isolation. In conclusion, this study makes a valuable contribution to the services industry by meticulously assessing the efficacy of deep learning classification techniques and exploring alternative strategies to effectively manage high-dimensional data challenges.

### III. RESEARCH METHODOLOGY

In scenarios involving time-varying features, various aggregation methods [23-24] have been explored alongside machine-learning classification techniques [26-27]. However,

these methods often fall short due to the requirement of having one observation per client in most classification techniques. This limitation becomes problematic when tracking the behavior of the same customer over time with time-varying features. Consequently, conventional classification methods struggle to effectively utilize this type of information.

In order to address our problem effectively, we have devised a structured approach that leverages a modified Variational Autoencoder (VAE). This method aims to uncover latent space attributes, overcome challenges in traditional autoencoders, and generate new features from complex datasets.

#### Step 1: Variational Autoencoder (VAE) Setup

We begin by setting up a Variational Autoencoder (VAE), a powerful tool known for its ability to uncover latent space attributes in data. The VAE comprises two essential components: an encoder and a decoder.

#### Step 2: Encoder and Decoder Functions

The encoder processes input data samples and maps them to latent variables. This encoder is instrumental in generating meaningful latent features. On the other hand, the decoder strives to replicate the input data using the learned latent variable distribution.

#### Step 3: Leveraging Latent Space

Latent variables are relatively low-dimensional representations of the input data, which contrasts with the high-dimensional input and reconstructed data. This approach is built on the idea that data is generated by the model  $P(x|z)$ .

#### Step 4: SV-VAE Architecture

Our proposed methodology incorporates four major blocks within the Space Vector Variational Autoencoder (SV-VAE): Encoder, Latent Distribution, KL Divergence, and Decoder. The SV-VAE leverages posterior distribution for data sampling and applies an empirical rule to reduce noise and approximate data points.

#### Step 5: Training the Model

Training involves optimizing two key loss functions: The KL divergence loss, which regularizes the learned latent distribution against a prior distribution, and the reconstruction loss, which ensures fidelity between decoded samples and original inputs.

#### Step 6: Deep Neural Network

Compressed features obtained from the SV-VAE are fed into a deep neural network with layers like pooling, dropout, ReLU, and a sigmoid layer. The output of this CNN flows into the decoder for data reconstruction.

#### Step 7: Evaluation

Model evaluation is performed through the assessment of SV-VAE loss, including the KL divergence loss function. Model predictions are evaluated for accuracy, F1-score, and precision. Hyperparameter tuning is carried out to enhance model accuracy.

#### Step 8: Dual Loss Optimization

Our SV-VAE model optimizes two crucial loss functions: The reconstruction loss, ensuring alignment with original dataset images, and the KL-divergence loss, quantifying the divergence from a standard normal distribution. This dual loss optimization ensures the model's effectiveness in capturing latent features.

The proposed model incorporates a modified Variational Autoencoder to uncover latent space attributes. VAE's ability to generate data across the entire space addresses the challenge of non-regularized latent space in traditional autoencoders. Within the VAE framework, an encoder module transforms the input sample  $x$  into a latent space representation  $x'$ . Variational autoencoders are particularly well-suited for generating new features from complex datasets [28].

The VAE consists of two core components: The encoder and the decoder. The encoder is a separate network that accepts samples from the data  $\{x_i\}_{i=1}^N$  and attempts to map them to the latent variables  $z$ . The decoder, on the other hand, attempts to replicate the input  $\{\hat{x}_i\}_{i=1}^N$  using the learned distribution  $z$ . Input  $x$  and reconstructed data samples  $\hat{x}$  are in high dimensional space, however, latent variable  $z$  is relatively low dimensional. The foundation of the variational autoencoder rests on the notion that data is generated by the model  $P(x|z)$ .

As illustrated in Fig. 1, the proposed methodology comprises four major blocks within the SV-VAE: Encoder, Latent Distribution, KL Divergence, and Decoder. The space vector variational autoencoder samples the data based on the posterior distribution. To remove noise and approximate data points, an empirical rule is applied. The encoder block plays a crucial role in generating latent features from the normal distribution data, emphasizing mean and standard deviation.

Training the model involves optimizing two loss functions: the KL divergence between the learned latent distribution and the prior distribution, which acts as a regularization term, and the reconstruction loss, which enforces fidelity between the decoded samples and the original inputs.

The compressed features from the SV-VAE are fed into a deep neural network that includes layers like pooling, dropout, ReLU, and a sigmoid layer. The output from the CNN flows to the decoder for data reconstruction. The evaluation of SV-VAE loss is accomplished through the KL divergence loss function. Similarly, the model's predictions are assessed for accuracy, f1-score, and precision. Hyperparameter tuning is conducted to enhance model accuracy.

The proposed SV-VAE model optimizes two key loss functions: reconstruction loss, which ensures that the decoder's output aligns with the original dataset images, and KL-divergence loss, which quantifies the divergence between the latent vector and a unit normal distribution. This divergence measurement ensures that the latent variables closely adhere to a standard normal distribution.

Martins [14] conducted a study revealing that the accuracy of Long Short-Term Memory (LSTM) models equipped with time series attributes is comparable to an approach that integrates this pertinent data using the mean and a random forest technique. This research outcome contributes to synthesizing insights from various studies.

Numerous research endeavors have underscored the efficacy of combining Convolutional Neural Networks (CNN) with Long Short-Term Memory (LSTM) models to enhance performance across diverse tasks. For example, in a study by Tan et al., [15], a fusion of CNN and LSTM was employed to forecast user-intended actions, demonstrating that this integration surpasses the individual performance of the two models and traditional machine learning methods. Furthermore, another investigation employed encoded sequential data, such as images and videos, wherein the utilization of CNN yielded superior outcomes compared to gradient boosting and random forest techniques [16].

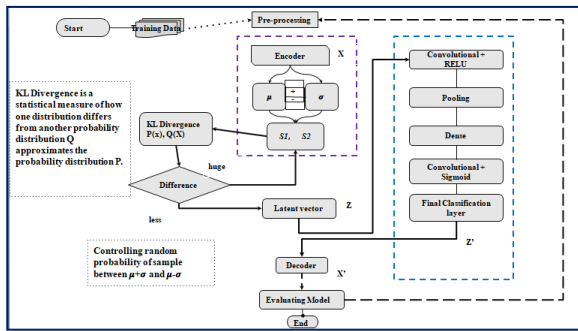


Fig. 1. Proposed SV-VAE model.

The sampling scenarios in VAE are to map the input to a distribution instead of mapping the input to a fixed vector.

$$x = \text{sample}(N(\mu, \sigma^2))$$

The modified approach uses mean and standard deviations to approximate the distribution of data,

$$x = (\mu + 2\sigma) + (-2\sigma) \text{sample}(N(0,1))$$

#### IV. EXPERIMENTAL SETUP

##### A. Dataset

This study utilizes six distinct publicly available datasets from diverse domains, sourced from repositories like Kaggle and UCI. The datasets encompass a range of data types, including discrete, continuous, and categorical values. The dataset sizes vary, with a minimum of 954 records from the Tour & Travels domain to a maximum of 15,000 records from the Music streaming subscriptions domain. Additional insights regarding the dataset characteristics and the specifics of the training-test split are provided in Table II.

TABLE II. DETAILED CHURN DATASET

Domain	Number of records			No of attributes	Train Set	Test Set
	Churn	Non Churn	Total			
Bank	32000	68000	10000	14	80000	20000
Telecom – fixed line	2000	5000	70000	19	5600	1400
Employee churn	270	1200	1470	35	1176	294
Online subscription	15000	135000	150000	30	120000	30000
Tour & Travels	109	845	954	7	763	191
Telecom – mobile	869	2281	3150	13	2520	630

##### B. Hardware and Software

The study was conducted on an Ubuntu 20.04 LTS operating system, employing an i9 12th-generation processor coupled with 16GB RAM and a 1TB HDD. The implementation process was carried out using a Jupiter Notebook in Python v3.10.0. For the implementation, a suite of Python libraries was utilized, encompassing NumPy, Pandas, Seaborn, Sklearn, Keras, TensorFlow v2.0, and Matplotlib. These libraries played pivotal roles in both data pre-processing and modeling stages, contributing to the overall analysis.

##### C. Pre-Processing

Data pre-processing is a fundamental phase in the workflow of every machine learning engineer. This stage encompasses a range of essential steps aimed at refining the dataset for optimal analysis. Imputation of missing values, type conversion, duplicate removal, cleansing, normalization, and transformation are key procedures frequently applied in this phase. For addressing missing values, diverse strategies can be employed, such as statistical methods like mean, median, or even leveraging regression models to predict and fill in the absent values. Data cleansing, on the other hand, entails eliminating noisy data through techniques like binning, regression, and clustering. Once the crucial pre-processing steps are completed, a thorough analysis of the attributes follows, often leading to the creation of new features. The process of attribute selection involves assessing the correlation between variables and selecting the appropriate number of attributes that contribute most effectively to the analysis. To prepare the data for subsequent modeling, it is transformed into a structured format, typically in the form of two-dimensional arrays. These arrays are then divided into training and testing sets. The training data, which constitutes the input for model training, is carefully configured to enable accurate analysis and prediction.

##### D. Hyperparameter

The model is fine-tuned through the manipulation of hyperparameters, which play a crucial role in enhancing the algorithm's performance. These hyperparameters encompass attributes such as batch size, optimizer, number of epochs, learning rate, dropout rate, and random initialization. By carefully adjusting these parameters, the algorithm can be optimized to yield a more generalized and accurate model. Batch size, a vital hyperparameter in gradient descent, determines the number of training data instances utilized in

each iteration. It governs the update of internal model parameters before proceeding to the subsequent iteration. The epoch parameter controls the iteration count for data feeding into the model. The learning rate hyperparameter adjusts the step size of weight adjustments during each epoch, critically influencing the optimization process. To guard against overfitting, a common challenge in model training, the dropout hyperparameter is introduced. This mechanism randomly omits a portion of neurons during training, preventing the model from becoming overly tailored to the training data. This practice enhances the model's capacity to generalize to unseen data.

In the context of this study, these hyperparameters are strategically manipulated to regulate the network's behavior during the training phase, ultimately contributing to the development of a more robust and efficient model.

### V. RESULT AND DISCUSSION

The model's performance was assessed using essential metrics such as precision, recall, accuracy, and F1-score. Depending on the specific business context, the choice between prioritizing precision or recall was determined to gauge the effectiveness of the churn model. Each of these metrics is mathematically derived to provide a comprehensive understanding of the model's classification capabilities. These quantitative assessments serve as valuable tools for objectively evaluating the model's performance, catering to different business needs and objectives. Mathematically each of these measurements is derived by,

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

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

$$\text{Accuracy} = \frac{TP + TN}{\text{Total}}$$

$$F1 - \text{Score} = 2 * \left( \frac{\text{Sensitivity} * \text{Specificity}}{\text{Sensitivity} + \text{Specificity}} \right)$$

TP = True Positive                      FP = False Positive

FN = False Negative                    TN = True Negative

TABLE III. SV-VAE WITH IMPACT DROPOUT OVER ACCURACY

Dropout rate	0.1	0.25	0.5	0.75	0.9
Accuracy					
Telecom - fixed line	94	94.25	91.0	89.45	81.6
Telecom	96.2	96.15	90.4	85.6	78.0
Tour and travel	97.8	98.1	88.3	80.2	72.6
Banking	93.1	90.5	85.26	79.36	72.0
Music online subscription model	95.8	96.3	87.1	80.1	74.0
Employee retention	98.1	98.2	90.0	85.0	74.5

To investigate the behavior of the proposed model, a range of dropout values was experimented with (see Table III). It was observed that dropout values between 0.1 and 0.25 yielded

optimal results, striking a balance between preventing overfitting and retaining useful information and Table IV shows the effect of different learning rates and average accuracy.

TABLE IV. EFFECT OF DIFFERENT LEARNING RATES AND AVG. ACCURACY

Learning rate	1	0.5	0.1	0.01	0.001	0.0001
Avg. Accuracy %	68.3	72.7	83.1	85.6	93.25	90.2

Table V displays the confusion matrix, offering a detailed breakdown of instances in which non-churn data is correctly classified as such (True Positives - TP) and churn data is accurately identified as churn (True Negatives - TN). In the context of a churn predictive model, the primary goal is the precise identification of churn users. This matrix provides a comprehensive assessment of the model's performance, shedding light on both accurate and erroneous classifications. Consequently, it informs the calculation of various evaluation metrics utilized in the analysis. In Fig. 2, a recall comparison between the proposed model and standard autoencoders is presented, highlighting the superior recall performance of the proposed model.

TABLE V. MODEL EVALUATION WITH CONFUSION MATRIX - ONLINE MUSIC STREAMING SUBSCRIPTION DATASET

	Actual Churn	Actual Not Churn
Predicted Churn	<b>4300</b>	<b>300</b>
Predicted Not Churn	<b>200</b>	<b>145000</b>
Total Records	<b>150000</b>	
Total Not Churn	<b>148000</b>	
Total Churn	<b>4500</b>	
Precision	<b>0.9712</b>	
Recall	<b>0.9966</b>	
Accuracy	<b>0.9680</b>	

The evaluation of recall enhancement not only underscores the technical progress achieved with the SV-VAE model but also carries profound implications for churn analysis. The notable uptick in the average recall, approximately 4.38%, signifies a substantial boost in the model's capability to accurately detect instances of churn. Within the realm of churn analysis, recall serves as a pivotal metric, quantifying the model's proficiency in capturing genuine churn occurrences among the overall churn cases. This enhancement directly translates into a more potent identification of customers at risk of churning, thereby equipping businesses with the proactive means to intervene and retain these valuable customers.

Table VI presents a comparative analysis of churn prediction accuracy using different types of autoencoders across various domains. Notably, the SV-VAE model consistently stands out, demonstrating superior accuracy across multiple sectors. The remarkable improvement of 5.01% in average accuracy underscores the model's enhanced ability to make precise classifications, distinguishing between churn and non-churn instances with greater accuracy. In churn analysis, accuracy is a vital metric that quantifies the overall correctness of the model's predictions. The improved accuracy ensures that the decisions based on the model's predictions are more

reliable, leading to optimized resource allocation for customer retention efforts and yielding better business outcomes. In both instances, these improvements in recall and accuracy substantiate the efficacy of the SV-VAE model in the realm of churn analysis. By accurately identifying potential churners and enhancing overall classification precision, the SV-VAE model enables businesses to devise more targeted and effective strategies to mitigate customer churn. This not only contributes to retaining valuable customers but also optimizes resource allocation and strategic decision-making, ultimately bolstering the competitive edge of businesses in the market.

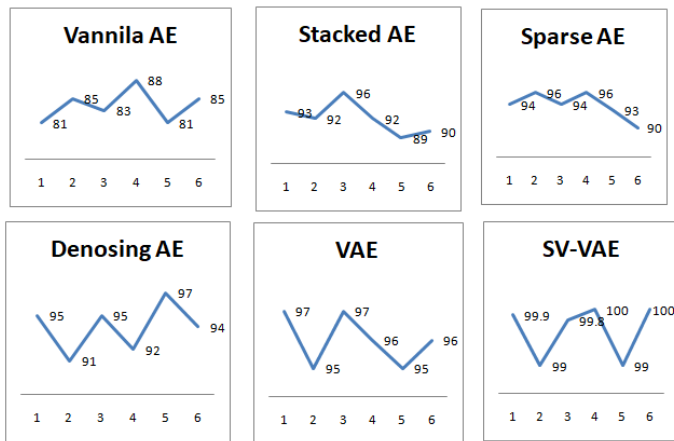


Fig. 2. Recall of the different types of autoencoders with proposed SV-VAE.

TABLE VI. CHURN PREDICTION ACCURACY OVER VARIOUS DOMAINS COMPARED AGAINST DIFFERENT TYPES OF AUTOENCODERS

Model/Dataset	Vanilla AE	Stacked AE	Sparse AE	Denoising AE	VAE	SV-VAE
Telecom Mobile	81	93	94	95	97	99.9
Bank	85	92	96	91	95	99
Music streaming	83	96	94	95	97	99.8
Employee	88	92	96	92	96	100
Telecom – Fixed line	81	89	93	97	95	99
Tour & Travels	85	90	90	94	96	100

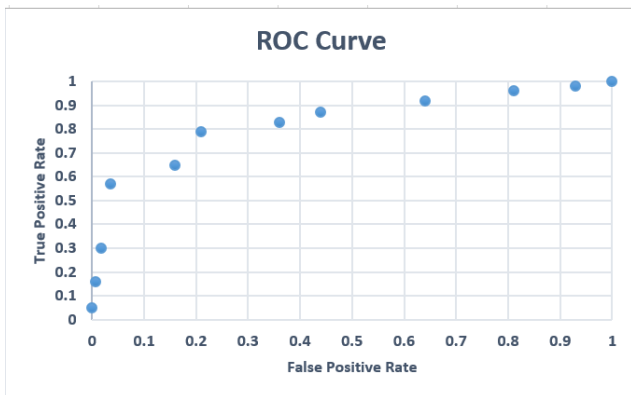


Fig. 3. ROC Curve of the proposed SV-VAE.

A ROC (Receiver Operating Characteristic) curve is a graphical tool that shows how the True Positive Rate (TPR) and False Positive Rate (FPR) change when we adjust the threshold for classifying data points as either positive or negative.

$$FPR = \frac{\text{False Positives}}{\text{False Positives} + \text{True Negatives}}$$

$$TPR = \frac{\text{True Positives}}{\text{False Positives} + \text{True Negatives}}$$

By adjusting this threshold, one can observe the variation in TPR (True Positive Rate) and FPR (False Positive Rate) values. Typically, as the threshold decreases, TPR increases, but FPR also rises. The ROC (Receiver Operating Characteristic) curve provides a visual representation of this trade-off and serves as a valuable tool for assessing a model's performance across various thresholds. The AUC (Area Under the Curve) is a quantitative metric that summarizes the model's overall performance over all conceivable thresholds. It quantifies the model's capacity to differentiate between positive and negative instances. In the case of the proposed model, the AUC value stands at 92.45. A greater AUC score denotes enhanced discriminatory power, with a value of 1 denoting a model that operates perfectly, and 0.5 indicating a model that merely makes random guesses. Fig. 3 graphically presents the ROC curve of the proposed model, offering a clear visual representation of its discriminatory power.

## VI. CONCLUSION AND FUTURE WORK

The proposed hybrid model, known as the Space Vector Variational Autoencoder with Convolutional Neural Networks (SV-VAE with CNN), represents a powerful solution for churn prediction tailored to the unique characteristics of churn data. The proposed approach represents a significant departure from prevailing systems in several key aspects. While contemporary systems often rely on traditional machine learning techniques and struggle to effectively utilize time-varying features, this method harnesses the power of a modified Variational Autoencoder (VAE) to unlock latent data attributes. This approach offers several distinctive advantages. It excels in the effective handling of time-varying features. Unlike conventional systems that face limitations when tracking the behavior of the same customer over time with time-varying features, this approach excels in this regard. By leveraging a VAE, it can capture the dynamic nature of features and generate latent representations that encapsulate temporal patterns. Additionally, it effectively reduces data dimensionality through the VAE's latent space, enabling more efficient analysis and modeling. In comparison to standard autoencoders, this approach incorporates Space Vector Variational Autoencoder (SV-VAE) architecture, enabling better discrimination and noise reduction, contributing to more accurate predictions. Moreover, while some systems focus solely on reconstruction loss, this approach optimizes two critical loss functions: reconstruction loss and KL-divergence loss, ensuring the effective capture of latent features while adhering to a standard normal distribution.

This research also opens avenues for generalization to more intricate scenarios and challenges.

**Multi-Modal Data Integration:** The approach, rooted in the VAE framework, can readily adapt to scenarios involving multi-modal data sources. By extending the encoder and decoder components, it can incorporate various data types and establish a more comprehensive understanding of complex cases.

**Temporal Sequence Modeling:** While addressing time-varying features, there is potential to explore more advanced temporal sequence modeling techniques, such as incorporating recurrent neural networks (RNNs) or attention mechanisms to capture intricate temporal dependencies.

**Transfer Learning and Scalability:** As the foundation of this approach lies in feature extraction and latent space representation, it is well-suited for transfer learning, allowing for the application of knowledge gained from one domain to another. Additionally, this methodology can be scaled to accommodate larger datasets and more extensive feature sets by leveraging distributed computing and parallel processing, extending its applicability to handle big data scenarios.

In conclusion, the proposed approach not only distinguishes itself from existing systems but also paves the way for broader applications in complex cases. These differences and potential generalization pathways are discussed here to provide a more comprehensive view of the research's contributions and future possibilities.

#### REFERENCES

- [1] Gerpott TJ, Rams W, Schindler A. Customer retention, loyalty, and satisfaction in the German mobile cellular telecommunications market. *Telecommun Policy*. 2001;25:249–69.
- [2] Wei CP, Chiu IT. Turning telecommunications call details to churn prediction: a data mining approach. *Expert Syst Appl*. 2002;23(2):103–12.
- [3] Qureshii SA, Rehman AS, Qamar AM, Kamal A, Rehman A. Telecommunication subscribers' churn prediction model using machine learning. In: Eighth international conference on digital information management. 2013. p. 131–6.
- [4] Ascarza E, Iyengar R, Schleicher M. The perils of proactive churn prevention using plan recommendations: evidence from a field experiment. *J Market Res*. 2016;53(1):46–60.
- [5] Bott. Predicting customer churn in telecom industry using multilayer perceptron neural networks: modeling and analysis. *Igarss*. 2014;11(1)
- [6] Umayaparvathi V, Iyakutti K. A survey on customer churn prediction in telecom industry: datasets, methods, and metrics. *Int Res J Eng Technol*. 2016;3(4):1065–70.
- [7] Yu W, Jutla DN, Sivakumar SC. A churn-strategy alignment model for managers in mobile telecom. In: Communication networks and services research conference, vol. 3. 2005. p. 48–53.
- [8] Mena, C.G., Caigny, A.D., Coussemont, K., Bock, K.W., & Lessmann, S. (2019). Churn Prediction with Sequential Data and Deep Neural Networks. A Comparative Analysis. *ArXiv*, abs/1909.11114.
- [9] De Caigny, A., K. Coussemont, K. W. D. Bock, and S. Lessmann (2019): "Incorporating textual information in customer churn prediction models based on a convolutional neural network," *International Journal of Forecasting*.
- [10] Allam, Swetha. (2019). Churn Prediction using Attention Based Autoencoder Network. *International Journal of Advanced Trends in Computer Science and Engineering*. 8. 725-730.
- [11] Brandusoiu I, Todorean G, Ha B. Methods for churn prediction in the prepaid mobile telecommunications industry. In: International conference on communications. 2016. p. 97–100.
- [12] Wei, C.-P. and I.-T. Chiu (2002): "Turning telecommunications call details to churn prediction: a data mining approach," *Expert Systems with Applications*, 23, 103 – 112
- [13] Song, G., D. Yang, L. Wu, T. Wang, and S. Tang (2006): "A Mixed Process Neural Network and its Application to Churn Prediction in Mobile Communications," in Sixth IEEE International Conference on Data Mining - Workshops (ICDMW'06), 798–802.
- [14] Martins, H. (2017): "Predicting user churn on streaming services using recurrent neural networks,"
- [15] Tan, F., Z. Wei, J. He, X. Wu, B. Peng, H. Liu, and Z. Yan (2018): "A, Blended Deep Learning Approach for Predicting User Intended Actions," 2018 IEEE International Conference on Data Mining (ICDM), 487–496.
- [16] Zaratiegui, J., A. Montoro, and F. Castanedo (2015): "Performing Highly Accurate Predictions Through Convolutional Networks for Actual Telecommunication Challenges," *CoRR*, abs/1511.04906.
- [17] Nasebah Almufadi, Ali Mustafa Qamar, Rehan Ullah Khan, Mohamed Tahar Ben Othman, Deep Learning-based Churn Prediction of Telecom Subscribers, *International Journal of Engineering Research and Technology*. ISSN 0974-3154, Volume 12, Number 12 (2019), pp. 2743-2748
- [18] Joolfoo, Muhammad. (2020). Customer Churn Prediction in Telecom Using Machine Learning in Big Data Platform.
- [19] Domingos, Edvaldo & Ojeme, Blessing & Daramola, Olawande. (2021). Experimental Analysis of Hyperparameters for Deep Learning-Based Churn Prediction in the Banking Sector. *Computation*. 9.0.3390/computation9030034.
- [20] Ammar A.Q. Ahmed, D. Maheswari, Churn prediction on huge telecom data using hybrid firefly based classification, *Egyptian Informatics Journal*, Volume 18, Issue 3,2017,Pages 215-220,ISSN 1110-8665,https://doi.org/10.1016/j.eij.2017.02.002.
- [21] Wang, Li and Chen, Chaochao and Zhou, Jun and Li, Xiaolong(2018) Time-sensitive Customer Churn Prediction based on PU Learning, *arXiv*, 10.48550/ARXIV.1802.09788
- [22] Umayaparvathi, V. & Iyakutti, K.. (2012). Applications of Data Mining Techniques in Telecom Churn Prediction. *International Journal of Computer Applications*. 42. 5-9. 10.5120/5814-8122.
- [23] Wangperawong, Ardit & Brun, Cyrille & Laudy, Olav & Pavasuthipaisit, Rujikorn. (2016). Churn analysis using deep convolutional neural networks and autoencoders.
- [24] Kristensen, Jeppe & Burelli, Paolo. (2019). Combining Sequential and Aggregated Data for Churn Prediction in Casual Freemium Games. 1-8. 10.1109/CIG.2019.8848106.
- [25] Prosvetov, Artem. (2018). The comparison of autoencoder architectures in improving of prediction models. *Journal of Physics: Conference Series*. 1117. 012006. 10.1088/1742-6596/1117/1/012006.
- [26] C. Wei and I. Chiu, Turning telecommunication call details to churn prediction: a data mining Approach expert System with applications, 2002.
- [27] Y. Liu, Z. Xu, J. Yang, L. Wang, C. Song and K. Chen, "A Novel Meta-Heuristic-Based Sequential Forward Feature Selection Approach for Anomaly Detection Systems," 2016 International Conference on Network and Information Systems for Computers (ICNISC), 2016, pp. 218-227, doi: 10.1109/ICNISC.2016.056.
- [28] Singh A, Ogunfunmi T. An Overview of Variational Autoencoders for Source Separation, Finance, and Bio-Signal Applications. *Entropy (Basel)*. 2021 Dec 28;24(1):55. doi: 10.3390/e24010055. PMID: 35052081; PMCID: PMC8774760.



# A New Method for Classifying Intracerebral Hemorrhage (ICH) Based on Diffusion Weighted – Magnetic Resonance Imaging (DW-MRI)

Andi Kurniawan Nugroho, Jajang Edi Priyanto, Dinar Mutiara Kusumo Nugraheni

Electrical Engineering Department, Universitas Semarang, Semarang, Indonesia

Public Health Department, Institut Kesehatan Indonesia, Jakarta, Indonesia

Informatic Department-Faculty of Science and Mathematics, Universitas Diponegoro, Semarang, Indonesia

**Abstract**—Stroke is a condition where the blood supply to the brain is cut off. This occurs due to the rupture of blood vessels in the intracerebral area or Intracerebral Hemorrhage (ICH). Examination by health workers is generally carried out to get an overview of the part of the brain of a patient who has had a stroke. The weakness in diagnosing this disease is that deeper knowledge is needed to classify the type of stroke, especially ICH. This study aims to use the Modified Layers Convolutional Neural Network (ML-CNN) method to classify ICH stroke images based on Diffusion-Weighted (DW) MRI. The data used in this study is a DWI stroke MRI image dataset of 3,484 images. The data consists of 1,742 normal and ICH images validated by a radiologist. Because the data used is relatively small and takes into account the computational time, Stochastic Gradient Descent (SGD) is used. This study compares the basic CNN model scenario with the addition of layers to the original CNN model to produce the highest accuracy value. Furthermore, each model is cross-validated with a different k to produce performance in each model as well as changes to batch size and epoch and comparison with machine learning models such as SVM, Random Forest, Extra Trees, and kNN. The results showed that the smaller the number of batch sizes, the higher the accuracy value and the number of epochs, the higher the accuracy value of 99.86%. Then, four machine learning methods with accuracy, sensitivity, and specificity below 90% are all compared to CNN2. As a summary of this research, the proposed CNN modification works better than the four machine learning models in classifying stroke images.

**Keywords**—Batch size; Epoch; ML-CNN; SGD; Stroke

## I. INTRODUCTION

There are more than 3.4 million new intracerebral hemorrhages each year. Globally, more than 28% of all stroke events are intracerebral hemorrhage. Annually, more than 23% of all intracerebral hemorrhages occur in persons aged 15-49 years [1]. Stroke is a fairly serious problem because stroke is a medical emergency that can threaten disability and death in patients if it is not handled quickly and appropriately. In the diagnosis of stroke, neurological imaging always plays an important role. Most stroke patients carry out examinations using Computerized Tomography Scanning (CT scan) radiology modalities. However, the CT scan image results for each patient vary according to the time interval that has passed since the stroke. Therefore, a radiologist plays a major role in

determining the management of stroke patients whether further imaging is carried out using Magnetic Resonance Imaging (MRI) to find out how severe the cell damage is in the brain, so that knowledge of the patient's radiological image will determine the treatment to be undertaken by the patient [2].

Stroke causes a reduction or stoppage of blood flow carrying oxygen which results in the death of brain cells. Based on the cause, stroke is divided into two, namely ischemic stroke where the blood supply stops flowing to the brain due to a blockage and hemorrhagic stroke where bleeding occurs in the brain tissue [3]. It is important to receive the correct diagnosis before stroke treatment begins, because treatment for stroke differs according to the type of stroke. If the patient fails to receive prompt and appropriate treatment, a stroke will have serious consequences and cause permanent damage to the brain and even death to the patient. Treatment and diagnosis of stroke is carried out by clinical examination, and then followed by examining radiologist modalities such as CT scan and MRI [4].

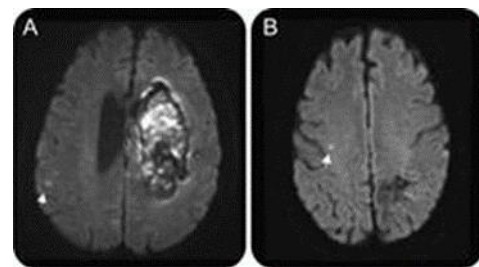


Fig. 1. DWI lesions in the acute and nonacute time periods.

Fig. 1 shown A visible parietal DWI lesion one day after a contralateral basal ganglia hemorrhage in a 66-year-old man (A) and a frontal DWI lesion two years after a contralateral parietal bleeding in a 69-year-old man (B). DWI = diffusion weighted imaging. Computational analysis of MRI images often aids physicians in diagnosis and helps reduce the subjectivity of diagnosis, and also provides higher accuracy for intensive care. Modern applications of artificial intelligence are designed to help humans solve various problems. CNN is a subcategory of deep learning development that is now widely used in neuroimaging [5]. Deep learning techniques and the use of CNN are often used to diagnose acute ischemic stroke. A popular topic in automated diagnostics is end-to-end system architecture. In this study, several CNN modifications are

proposed to find the best performance value in classifying ICH stroke and normal conditions so that doctors can quickly make the quickest diagnosis to determine the therapy given. This paper consists of five sections: the Section I contains an introduction; Section II contains related work; Section III contains the methodology used in the research; Section IV contains performance results from three CNN models, hyperparameters, and comparisons with machine learning models; and Section V contains conclusions and future research work.

## II. RELATED WORK

Recently, several important publications have presented implemented algorithms classifying brain stroke. In addition, the ischemic stroke lesion segmentation method and risk prediction are also applicable to stroke diagnosis [6]. Several studies have used common deep learning models such as Inception-V3 and EfficientNet-b0 to detect acute stroke using DW-MRI with an accuracy value of 86.3% [7]. A study related to the diagnosis and prediction of stroke by developing a detection system for only one type of stroke have detected early ischemia automatically using the Convolutional Neural Network (CNN) algorithm with 256 original images and augmented images, with the classification results obtaining an accuracy value of 90% [8]. Another study used the CNN algorithm with an open stroke dataset from [www.radiopaedia.org](http://www.radiopaedia.org) to classify patient data into three classes, namely normal, ischemic stroke, and hemorrhagic stroke through CT scan images.

Meanwhile, other researchers using the same dataset performed hyperparameter optimization in the Deep Learning algorithm to improve the accuracy of stroke diagnosis using CT scan image segmentation with the thresholding method and the binarization process. The implementation of the threshold method uses global binary thresholding and Otsu thresholding [9].

I.P. Kerta et al. [10] have segmented patient data to produce patient class labels and classify the results of grouping data to test the performance of the classification algorithm used. A total of 4,906 patient data used in this study were grouped using the K-Means method into several clusters, including two clusters, three clusters, four clusters, and five clusters, and the findings of these data groupings will be classified. The classification results produce the best accuracy value on the number of clusters tested, namely two clusters of 99.71%.

Jenna and Kumar [11] have performed a stroke classification using International Stroke trial data. The database includes patient information, patient history, hospital details, risk factors, and symptoms. Preprocessing is done to eliminate missing and inconsistent data. After preprocessing, 350 samples were taken in this work, with parameters sensitivity, specificity, accuracy, precision and F1 scores calculated to evaluate the performance of various kernel functions of the SVM classifier. The experimental results obtained the best precision in the kernel linear function with an accuracy value of 91%.

P. Govindarajan et al [12] presented a prototype for classifying strokes that combines text mining and machine

learning algorithms. At the data collection stage, patient data from 507 patients were collected from Sugam Multispecialty Hospital, Kumbakonam, Tamil Nadu, India. Processed data is fed into various machine learning algorithms such as artificial neural networks, Support Vector Machine (SVM), and random forest. Among these algorithms, the neural network trained with the Stochastic Gradient Descent algorithm outperforms other algorithms with a higher classification accuracy of 95%.

Y. Q. Zhang et al investigated the ability of a machine learning model based on MRI radiomic features (ML) to classify time since stroke onset (TSS), which could aid in stroke assessment and treatment options. This study involved 84 patients with acute ischemic stroke. Segmentation of the infarct area is made manually with 3D-slicer software. A total of 4312 radiomic features from each image sequence were captured and used in six machine learning models to estimate stroke onset time for binary classification ( $\leq 4.5$  hours). Receiver-Operating Characteristic (ROC) curves and other parameters are calculated to evaluate the performance of the model in the training and test groups. Twelve radiomic results and six clinical features were selected to construct the ML model for TSS classification. The deep learning model-based DWI/ADC radiomic feature showed the best for binary TSS classification in the independent test group, with AUC 0.754, accuracy 0.788, sensitivity 0.952, specificity 0.500, positive predictive value 0.769, and negative predictive value 0.857, respectively [13].

Our contributions to this study are as follows:

1) We propose a classification of stroke intracerebral hemorrhage (ICH) using MRI images with modifications to the addition of a convolution layer to the simple CNN model with hyperparameter tuning, such as changes in epoch, batch size, and the use of k- fold validation in knowing performance values from a limited number of datasets.

2) We have compared the performance of Modified Layer CNN (ML-CNN) with machine learning models to produce a good method for classifying DW- MRI ICH stroke images with normal images.

Several CNN models were analyzed using several optimal methods to compare the performance of various machine learning algorithm approaches using four parameters, namely accuracy, precision, recall, f1- measure, and k-fold cross-validation to optimize performance, resulting in high prediction accuracy.

## III. METHODOLOGY

### A. Data Collection

The stages of data collection in this study were to collect image data from DW-MRI of the patient's brain consisting of Intracerebral Hemorrhage (ICH) stroke image data and normal image data. DW-MRI image data comes from Gatot Subroto Hospital Jakarta which was taken during the January-May 2019 period and came from 430 patients with ICH stroke indications. The image taken has a slice thickness of 5.0 mm, the distance between slices is 6.5 mm, the pixel spacing is 0.7 mm and the original image size is 320 pixels x 320 pixels.

The data that has been collected is labeled to distinguish between ICH stroke data and normal data by radiologists. To eliminate noise, a filter is performed and to add data to prevent overfitting, augmentation is carried out in real time. Images were entered for each CNN model and the values for accuracy, precision, recall, f1-score, and specificity were calculated. Next, the collected data is compared with machine learning models to find out how high the performance of each of these models is. The research method is shown in Fig. 2.

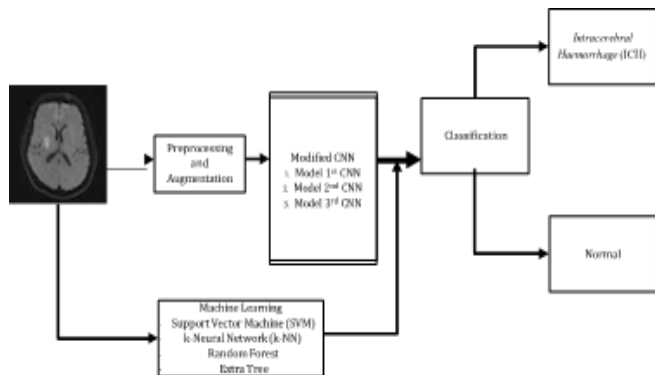


Fig. 2. Research method.

**B. CNN Models**

This research was conducted using three variations of the CNN architecture to obtain the most appropriate architecture in detecting and differentiating the presence of intracerebral hemorrhage (ICH) stroke and normal brain images. Computations are performed using NVIDIA GeForce GTX 1650 GPU RAM 16 GB 2600 MHz DDR4 to shorten the compilation time of the CNN program.

The architectural design is presented in Table I. Fig. 3 shows the basic architecture of the CNN method (CNN1). Fig. 4 (CNN2) shows an additional development of the convolution layer from CNN1. Fig. 5 (CNN3) is the development of the CNN2 model by adding each layer and its activation function to get the classification accuracy value. Comparisons can be made between these designs individually (CNN1, CNN2, CNN3) or between design groups to obtain the most appropriate type of design in classifying DW-MRI image data.

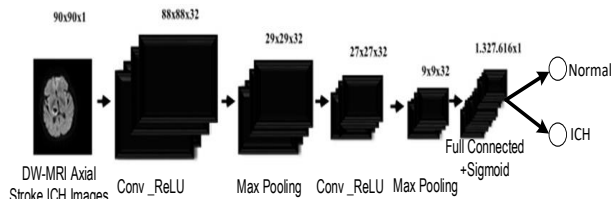


Fig. 3. CNN1 architecture visualization.

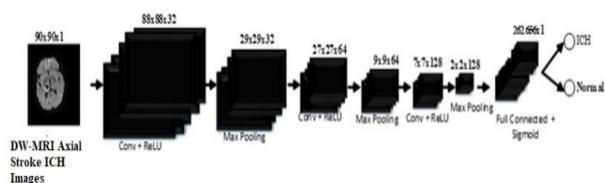


Fig. 4. CNN2 architecture visualization.

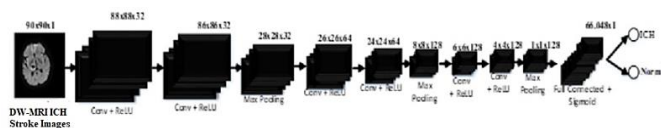


Fig. 5. CNN3 architecture visualization.

TABLE I. CNN CONFIGURATION TESTED

CNN1	CNN2	CNN3
Input Layer = 90x90x1	Input Layer = 90x90x1	Input Layer = 90x90x1
Conv. Layer (layer contains 32 filter of [3 3])	Conv. Layer (layer contains 32 filter of [3 3])	Conv. Layer (layer contains 32 filter of [3 3])
ReLU	ReLU	Conv. Layer (layer contains 32 filter of [3 3])
MaxPOOL (3x3, with stride [1 1])	MaxPOOL (3x3, with stride [1 1])	ReLU
Conv. Layer (layer contains 32 filter of [3 3])	Conv. Layer (layer contains 64 filter of [3 3])	MaxPOOL (3x3, with stride [1 1])
ReLU	ReLU	Conv. Layer (layer contains 64 filter of [3 3])
MaxPOOL (3x3, with stride [1 1])	MaxPOOL (3x3, with stride [1 1])	Conv. Layer (layer contains 64 filter of [3 3])
Dropout Layer (drop probability = 0.5)	Conv. Layer (layer contains 128 filter of [3 3])	ReLU
Full Connected Layer	ReLU	MaxPOOL (3x3, with stride [1 1])
Sigmoid Layer	MaxPOOL (3x3, with stride [1 1])	Conv. Layer (layer contains 128 filter of [3 3])
Classification Layer	Dropout Layer (drop probability = 0.5)	Conv. Layer (layer contains 128 filter of [3 3])
	Full Connected Layer	ReLU
	Sigmoid Layer	MaxPOOL (3x3, with stride [1 1])
	Classification Layer	Dropout Layer (drop probability = 0.5)
	Full Connected Layer	ReLU
	Sigmoid Layer	MaxPOOL (3x3, with stride [1 1])
	Classification Layer	Dropout Layer (drop probability = 0.5)
		Full Connected Layer
		Sigmoid Layer
		Classification Layer

**C. Augmentation Data**

Data Augmentation is a technique to increase the diversity of image data by performing basic transformations such as rotation, shared, horizontal flip, and zoom. According to Luis Perez et al. [14] by using this technique, the model can overcome the problem of overfitting and improve the accuracy of the CNN model. In this study, each class (ICH and Normal) used 430 data with 4 (four) geometric transformations, and generated data for each class of 1720 datasets per class. In each epoch, the model receives images with different transformations. This study applies real time data augmentation [15].

Configuration of data augmentation in this study is in the form of:

- 1) *Shared range* = 20 skews the image to a maximum angle of 20 degrees. The value used should not be too large because it will make the image flat.
- 2) *Rotation range* = 20 randomly rotates the image up to a maximum angle of 20 degrees.
- 3) *Horizontal flip* = True flips the image horizontally.
- 4) *Zoom range* = [0.75-1.0] randomly enlarges the image. A value of 0.75 means that the image is enlarged to 75%, while 1.0 is a normal image size.

Fig. 6 shows the real time augmentation results that are generated after running the CNN model training.

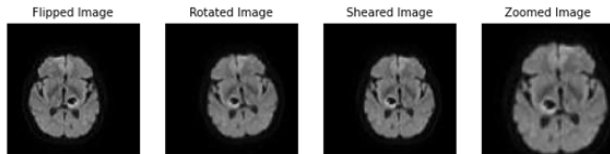


Fig. 6. DW-MRI image augmentation results.

#### D. K-Fold cross-validation

Evaluating machine learning models is very difficult. Usually, to divide the dataset into training and test sets it is necessary to use a training set to train the model and a test set to test the model. The next step is to evaluate the performance of the model based on the error matrix to determine the accuracy of the model. However, this method is not very reliable because the accuracy obtained for one test set can be very different from the accuracy obtained for different test sets. K-fold Cross-validation (CV) provides a solution to this problem by dividing the data into folds and ensuring that each fold is used as a test set at multiple CV points [16] [17].

This study uses 3,484 data. If using five-fold cross-validation, the data is divided into five folds of the same size where four folds will be used as training data and one fold is used as validation data. From a total of 3,484 ICH and normal image data, 2,787 data will be used at the training stage and 697 at the testing stage as test data. A total of 2,787 data used at the training stage will be divided into five (Each fold consists of 557 data), so that the amount of training data used is 2,230 data and validation data used is 557.

#### E. Performance Matrix for Classification

Three CNN models and machine learning algorithms were trained and evaluated by comparing four performance matrices such as: accuracy, precision, recall, and F1-score [16]:

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$F\ Score = 2x \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

FN, FP, TN and TP are False Negative, False Positive, True Negative, and True Positive.

## IV. RESULTS AND ANALYSIS

### A. CNN Model Performance Test

Comparisons are made by means of one parameter being a variable and the other parameters being assigned the same value for the three types of CNN architectures. Parameters compared were epoch, batch, and classification accuracy. Meanwhile, the other parameters are set to the same value, namely the number of filters is 32.64 and 128, with Stride 1 x 1, kernel 3 x 3, kernel pooling size 3 x 3, and using the SGD optimization function because it is stochastic. This means sampling random training data at each step, and then calculating gradients making it much faster because there is less data to manipulate at any one time. The image used as input from CNN is a rescaled grayscale image so that the size is 90 x 90 pixels.

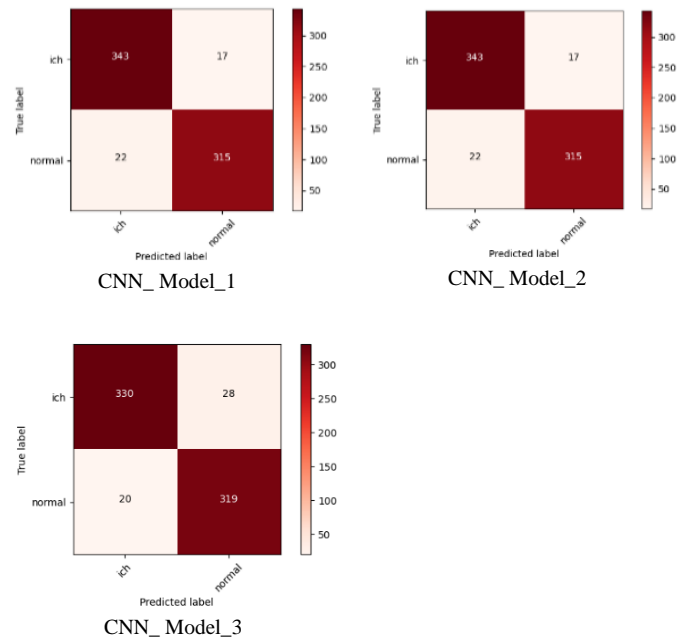


Fig. 7. Confusion matrix for classification of Intracerebral Hemorrhage (ICH) stroke and normal brain images using 3 modified CNN models.

Based on the results of the tests that have been carried out, a confusion matrix for Intracerebral Hemorrhage (ICH) and normal brain images can be made using three modified CNN models as presented in Fig. 7. The test results show that the confusion matrix classification of the CNN\_1 model and the CNN\_2 model with 343 ICH image values is correctly predicted and 17 ICH images were not correctly predicted. Likewise, 315 normal images were correctly predicted as normal images and two normal images were predicted incorrectly as normal images. On the other hand, model\_3 shows that 330 ICH images are correctly predicted as ICH images and 28 ICH images are predicted incorrectly as ICH images. Meanwhile, there are 319 normal images that can be predicted as ICH images, and only 20 ICH images that are predicted as ICH images.

Based on the performance results of the three CNN models shown in Table II, it can be seen that the accuracy values for

CNN model\_1 and CNN model\_2 have better performance than CNN\_3. This is because the more convolution layers are generated, the more map features will also be. Therefore, the system experiences overfitting in classifying the image dataset.

TABLE II. CNN PERFORMANCE ANALYSIS

CNN MODELS	Precision (%)	Recall (%)	Specificity (%)	F1_score (%)	Acc (%)
CNN1	95%	93%	95%	94%	94.40 %
CNN2	95%	93%	95%	94%	94.40 %
CNN3	92%	94%	92%	93%	94.11 %

The algorithm used to test the validity of the accuracy results is k = 3, 5, 7 and 10 Cross-validation. The dataset is divided according to the number of k-folds into five folds in which there are 1,720 datasets, and at each iteration one fold is taken as a testing dataset and the other is used as a training dataset. The selection of the dataset for testing is adjusted according to the iteration order and the fold order, namely the 1st iteration folds 1, the 2nd iteration folds 2, and so on. After each training is finished, testing is immediately carried out to find the predicted value, and then the level of accuracy is calculated on average. The test results are shown in Table III.

TABLE III. EFFECT OF CROSS-VALIDATION

CNN Models	Precision (%)	Recall (%)	Specificity (%)	F1_score (%)	Acc (%)
CV (n=3)					
CNN1	66%	79%	59%	72%	68.95%
CNN2	70%	93%	61%	80%	76.81%
CNN3	58%	61%	56%	60%	58.48%
CV (n=5)					
CNN1	86%	84%	86%	85%	85.20%
CNN2	89%	85%	89%	87%	87.00%
CNN3	72%	86%	67%	87%	76%
CV (n=7)					
CNN1	93%	76%	94%	83%	84.84%
CNN2	86%	91%	85%	89%	88.09%
CNN3	73%	90%	66%	80%	77.98%
CV (n=10)					
CNN1	86%	84%	86%	85%	84.84%
CNN2	86%	87%	85%	87%	86.28%
CNN3	70%	76%	68%	73%	71.74%

Table III shows that for each k-fold tested, the CNN2 model has the best performance value compared to the CNN1 and CNN3 models.

B. Effect of Batch Size

In this test, the epoch value is set at 30 with SGD optimization and a dropout value of 0.5. Usually, large batch sizes are used because they allow computational acceleration. If you use a small batch size, it will take a very long time. However, there is a price to pay behind the speed of computing. Batch sizes that are too large will produce less than optimal results. The larger the batch size, the less accurate the results will be [18].

The test results are shown in Table IV, V, and VI. It can be seen that batch size 8 has the best performance with the highest accuracy of the several variations in batch size values for the three models tested. The larger the batch size value, the lower the accuracy value of each CNN model.

TABLE IV. EFFECT OF BATCH SIZE MODEL CNN\_1

Number of Batch Size	Precision (%)	Recall (%)	Specificity (%)	F1_score (%)	Acc (%)
8	100%	100%	100%	100%	100.00 %
16	99%	96%	99%	98%	97.70%
32	94%	72%	95%	81%	83.00%
64	72%	78%	73%	75%	75.40%
128	69%	80%	64%	75%	72.02%
256	51%	56%	40%	54%	48.78%

TABLE V. EFFECT OF THE BATCH SIZE MODEL CNN\_2

Number of Batch Size	Precision (%)	Recall (%)	Specificity (%)	F1_score (%)	Acc (%)
8	100%	100%	100%	100%	99.86%
16	99%	97%	99%	98%	98.13%
32	72%	95%	61%	82%	78.48%
64	75%	85%	74%	77%	77.04%
128	69%	88%	60%	77%	74.03%
256	50%	63%	35%	56%	49.21%

TABLE VI. EFFECT OF BATCH SIZE MODEL CNN\_3

Number of Batch Size	Precision (%)	Recall (%)	Specificity (%)	F1_score (%)	Acc (%)
8	100%	98%	100%	99%	98.71%
16	100%	60%	100%	75%	80.06%
32	79%	53%	85%	64%	68.58%
64	71%	68%	72%	70%	70.16%
128	58%	99%	17%	73%	60.98%
256	51%	99%	5%	68%	52.65%

### C. Effect of Number of Epoch

The next design trial was carried out with a batch size of 32 with SGD optimization and a dropout value of 0.5, and the number of epochs varied. The test results are shown in Table VII. The greater the epoch, the higher the accuracy acquired, namely 99.86% accuracy at epoch 90, which is the greatest accuracy value in the ICH stroke dataset.

TABLE VII. DEGREE OF ACCURACY WHEN THE NUMBER OF EPOCHS VARIES

Number of Epochs	Precision (%)	Recall (%)	Specificity (%)	F1_score (%)	Acc (%)
30	73%	98%	63%	84%	80.77%
50	77%	100%	71%	87%	85.51%
70	99%	98%	99%	98%	98.13%
90	100%	100%	100%	100%	99.86%

### D. Comparison with Other Classification Methods

The following trials were conducted to compare the performance of the CNN2 Model design (as the best proven CNN design) with other classification methods, namely, SVM, Random Forest, Extra Trees, K Neighbors. For k-NN, the DW-MRI image dataset was tested using the k parameter of 3.

Meanwhile, for the SVM method, the dataset was tested using a linear kernel. Tests were carried out to classify the DW-MRI image dataset in the ICH or normal stroke class. The input data for the k-NN and SVM classifiers are the grayscale intensity values of each image pixel in the dataset.

The dataset used with all machine learning models uses 1720 normal images and 1720 ICH images with a training data and testing data ratio of 70:30. From the test results presented in Table VIII, it was found that the four machine learning methods (SVM, Random Forest, kNN, and Extra Trees) had poor performance, with levels of accuracy, sensitivity, and specificity all below 90%.

The CNN 2 model method produces the highest performance. This shows that the CNN 2 method can be implemented to classify DW-MRI images of stroke Intracerebral Hemorrhage (ICH) and normal brain images with good performance.

TABLE VIII. PERFORMANCE COMPARISON OF THE CNN2-SVM-KNN-RANDOM FOREST-EXTRA TREES METHODS ON ICH STROKE CLASSIFICATION

Method	Precision (%)	Recall (%)	Specificity (%)	F1_score (%)	Acc (%)
CNN2	95%	93%	9500%	94%	94.40%
SVM	70%	70%	6900%	70%	70%
Random Forest	62%	79%	2500%	56%	62%
Extra Tress	57%	77%	14%	47%	57%
K Neighbors (k=3)	36%	36%	41%	36%	36%

## V. CONCLUSIONS

From the results of the tests and analyzes that have been carried out, it can be concluded that the CNN algorithm can be used properly in the classification of DW-MRI images to distinguish stroke ICH images from normal DW-MRI images. In the first trial related to the CNN model performance test, the highest accuracy value was 94.40% for CNN1 and CNN2 compared to CNN3 because the more layers, the lower the accuracy value. To produce further performance on the CNN1 and CNN2 models, it was tested using the cross-validation method, the highest accuracy value was generated in the CNN 2 model for several variations of the number of k folds. Each CNN model was tested by changing the number of batch sizes. From the test results, the smaller the number of batch sizes (8), the higher the accuracy value and the number of epochs, the higher the number of epochs (90) the higher the accuracy value of 99.86%.

This study compares the basic CNN model scenario with the addition of layers to the original CNN model to produce the highest accuracy value. The dataset used is 1720 images for each class. Furthermore, each model is cross-validated with a different k to produce performance in each model as well as changes to batch size and epoch and comparison with machine learning models such as SVM, Random Forest, Extra Trees, and kNN. The results showed that the smaller the number of batch sizes, the higher the accuracy value and the number of epochs, the higher the number of epochs, the higher the accuracy value of 99.86%. Then, four machine learning methods with accuracy, sensitivity, and specificity below 90% are all compared to CNN2. The proposed CNN modification works better than the four machine learning models in classifying stroke images.

In the future, this study will be developed in terms of using CNN to classify 3D images so that classification classes can be multiplied. An example is not only to find out DW-MRI images of brain hemorrhage (ICH) or normal DW-MRI images, but can also find out blockages in several locations in the brain vessels with DW-MRI images.

## REFERENCES

- [1] W. S. O. (WSO), "Global Stroke Fact Sheet 2022," 2022. [Online]. Available: [https://www.worldstroke.org/assets/downloads/WSO\\_Global\\_Stroke\\_Fact\\_Sheet.pdf](https://www.worldstroke.org/assets/downloads/WSO_Global_Stroke_Fact_Sheet.pdf).
- [2] Y. Yueniwati, PENCITRAAN PADA STROKE. Malang, Indonesia: Universitas Brawijaya Press (UB Press), 2016.
- [3] C. I. L. Sam and A. N. , BN Mahasena Putera Awatara, DPG Purwa Samatra, "Penentuan Stroke Hemoragik dan Non-Hemoragik Memakai Skoring Stroke," Callosum Neurol., no. October, 2018, doi: 10.29342/cnj.v1i3.30.
- [4] M. A. Inamdar et al., "A Review on Computer Aided Diagnosis of Acute Brain Stroke," sensors, pp. 1-35, 2021.
- [5] Q. Bao et al., "MDAN: Mirror Difference Aware Network for Brain Stroke Lesion Segmentation," IEEE J. Biomed. Heal. Informatics, no. 21404050, 2021, doi: 10.1109/JBHI.2021.3113460.
- [6] B. Omarov, A. Tursynova, O. Postolache, K. Gamry, and A. Batyrbekov, "Modified UNet Model for Brain Stroke Lesion Segmentation on Computed Tomography Images," Comput. Mater. Contin., vol. 71, no. 3, pp. 4701-4716, 2022, doi: 10.32604/cmc.2022.020998.

- [7] K. Lee and D. Y. Chen, "Automatic detection and vascular territory classification of hyperacute staged ischemic stroke on diffusion weighted image using convolutional neural networks," *Eur. J. Radiol.*, 2022.
- [8] C. Chin and B. Lin, "An Automated Early Ischemic Stroke Detection System using CNN Deep Learning Algorithm," *IEEE 8th Int. Conf. Aware. Sci. Technol. (iCAST 2017)*, no. iCAST, pp. 368–372, 2017, doi: 10.1109/ICAwST.2017.8256481.
- [9] T. Badriyah, N. Sakinah, I. Syarif, and D. R. Syarif, "Segmentation Stroke Objects based on CT Scan Image using Thresholding Method," 2019.
- [10] I. P. Kerta, N. Kadek, D. Rusjyanthi, W. Siti, M. Binti, and M. Luthfi, "Classification of Stroke Using K-Means and Deep Learning Methods," *LONTAR Komput.*, vol. 13, no. 1, pp. 23–34, 2022.
- [11] Jeenar; Kumar, "Stroke Prediction Using SVM," in *2016 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)*, 2016, pp. 600–602.
- [12] P. Govindarajan, R. Kattur, and S. Amir, "Classification of stroke disease using machine learning algorithms," *Neural Comput. Appl.*, vol. 32, no. 3, pp. 817–828, 2020, doi: 10.1007/s00521-019-04041-y.
- [13] Y. Q. Zhang et al., "MRI radiomic features - based machine learning approach to classify ischemic stroke onset time," *J. Neurol.*, vol. 269, no. 1, pp. 350–360, 2022, doi: 10.1007/s00415-021-10638-y.
- [14] P. Aryasuta Wicaksana, "Pengenalan Pola Motif Kain Tenun Gringsing Menggunakan Metode Convolutional Neural Network Dengan Model Arsitektur," *Spektrum*, vol. 6, no. 3, pp. 159–168, 2019.
- [15] D. Mutiara, K. Nugraheni, A. K. Nugroho, D. Intan, K. Dewi, and B. Noranita, "Deca Convolutional Layer Neural Network ( DCL-NN ) Method for Categorizing Concrete Cracks in Heritage Building," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 1, pp. 722–730, 2023.
- [16] A. K. Nugroho and M. H. Purnomo, "Quad Convolutional Layers ( QCL ) CNN Approach for Classification of Brain Stroke in Diffusion Weighted ( DW ) - Magnetic Resonance Images ( MRI )," *Int. J. Intell. Eng. Syst.*, vol. 15, no. 1, pp. 414–427, 2022, doi: 10.22266/ijies2022.0228.38.
- [17] A. K. Nugroho, "Utilizing the Hepta Convolutional Layer Neural Network ( HCL-NN ) Based on a Multi Optimizer for the Classification of Brain Stroke MRI," *Int. J. Intell. Eng. Syst.*, vol. 15, no. 3, pp. 304–318, 2022, doi: 10.22266/ijies2022.0630.26.
- [18] N. Rochmawati, H. B. Hidayati, and Y. Yamasari, "Analisa Learning rate dan Batch size Pada Klasifikasi Covid Menggunakan Deep learning dengan Optimizer Adam," *J. Inf. Eng. Educ. Technol.*, vol. 05, pp. 44–48, 2021.

# Application Prototype for Inclusive Literacy for People with Reading Disabilities

Laberiano Andrade-Arenas<sup>1</sup>, Roberto Santiago Bellido-García<sup>2</sup>, Pedro Molina-Velarde<sup>3</sup>, Cesar Yactayo-Arias<sup>4</sup>

Facultad de Ciencias e Ingeniería, Universidad de Ciencias y Humanidades, Lima, Perú<sup>1</sup>

Departamento de Estudios Generales, Universidad César vallejo, Lima, Perú<sup>2</sup>

Facultad de Ingeniería, Universidad Tecnológica del Perú, Lima, Perú<sup>3</sup>

Departamento de Estudios Generales, Universidad Continental, Lima, Perú<sup>4</sup>

**Abstract**—This article details the process of creating a prototype mobile application that aims to promote inclusive literacy for people with reading disabilities. The goal of this application is to help people with reading difficulties to become more independent so that they can participate in society and take advantage of educational and employment opportunities that were previously unavailable to them. The methodology used in this work is Design Thinking as it is a user-centered creative approach to solving difficult challenges and addresses creativity, design and problem solving. The results obtained from the expert judgment based on Atlas TI 22 provide a valuable perspective on the viability and potential of these technological tools. The analysis of the results of the application prototype designs gives an encouraging picture of 85%. Similarly, 75% confirm that the app effectively complements inclusive literacy efforts, a significant achievement in line with the objective, and 70% appreciate the app's interaction with people with reading disabilities. Finally, a staggering 87% would gladly recommend the app, underscoring its valuable impact. In conclusion, the article discusses how mobile applications can help people with reading difficulties become more literate. The good reception of the prototype confirms the importance of technology in inclusive education and the value of this approach to improving the lives and education of this demographic.

**Keywords**—Atlas TI 22; inclusive literacy; mobile applications; reading disability; design thinking

## I. INTRODUCTION

The aim of this study was to report on the design of specialized reading materials, such as books with altered fonts and simplified text [1]. According to the findings, this type of content helped people with reading difficulties to read and comprehend what they read. The authors also discuss the advantages of cooperative and group learning for language acquisition [2]. The researchers found that when people with reading difficulties worked together in small groups, they were able to encourage and support each other as they learned.

However, despite advances in technology and education, considerable impediments remain in the way of literacy for people with reading difficulties [3]. Many potential elements come into play here, including cognitive, sensory, financial, and lack of specialized educational resources [4]. Symptoms can range from having trouble interpreting words to having trouble understanding what they read [5]. The impact has far-reaching effects, limiting social, employment and educational opportunities and prolonging marginalization.

Promoting inclusion and empowering people with reading difficulties requires action [6]. The study and creation of mobile applications tailored to inclusive literacy not only have the capacity to remove conventional barriers but also to reinvigorate learners [7]. With the results of this study, we hope to develop effective strategies to respond to the educational challenges faced by this population and give them access to resources that take into account their particular strengths and preferences [8]. The positive effects of promoting diversity and inclusion in the community are not limited to the individuals directly involved [9].

The importance of this article is to report on the creation of a prototype mobile application for inclusive literacy, aimed at users with reading difficulties. In the same way provide a welcoming and individualized classroom environment in which students can work on their reading comprehension problems [10]. The goal of this application is to help people with reading difficulties to become more independent so that they can participate in society and take advantage of educational and employment opportunities that were previously unavailable to them.

An innovative and potentially fruitful solution to the pedagogical difficulties faced by people with reading disabilities is presented: the development of mobile applications for inclusive literacy. This article delves into the global context, explains the problem, explains why this study is important, and sets the goal of creating a prototype mobile application that will help create a more just and egalitarian world.

The structure of the research is based on the following: Section II will present the literature review, Section III will present the methodology used in the research, Section IV will present the results, Section V will present the discussions and finally Section VI will present the conclusions and future work.

## II. LITERATURE REVIEW

For students with reading difficulties, inclusive literacy is a rapidly expanding area of study. The purpose of this literature review is to examine the various methods and techniques employed in inclusive literacy for individuals with reading difficulties. The education and literacy of this population will be examined along with research, programs, and practices that aim to improve their accessibility.



The authors [11] studied how mobile apps and screen readers, two examples of accessible technologies, can help people with reading difficulties become more literate. The results showed that text comprehension improved by using read-aloud features and by modifying the material. This group's reading comprehension and access to information were greatly enhanced by the use of technology.

The results of this study focus on the effectiveness of using flexible methods of teaching and reading. The authors [12] emphasize that students' reading comprehension and engagement increased when individualized tactics such as guided reading and the use of pictograms were introduced. These results underscore the need for personalized approaches to reading and writing instruction. The research also focused on the production of accessible literature for people with reading difficulties, such as simplified texts and audiobooks [13]. The results showed that the use of these modified materials increased interest in reading and improved comprehension. There is a broad consensus that the availability of literature in accessible formats is crucial to enable and encourage reading autonomy.

They discuss individuals with reading problems and the effects of teacher training in inclusive practices on their literacy. Also, the authors [14], teachers who received professional training were more adept at modifying lessons and providing students with individualized attention. As a result, the children's reading ability improved significantly, demonstrating the value of inclusive education. The study also analyzed the effectiveness of collaboration between teachers, speech-language pathologists and assistive technology specialists. Using a combination of methods from different fields, specialists were better able to meet the specific needs of individuals. The results underscored the need for inclusive literacy strategies.

Similarly, the authors [15] studied people with reading problems to see how increasing their literacy levels affected their ability to relate to others and feel self-confident. Those who made efforts to improve their skills felt more ownership of their lives and had easier access to resources. Because of its good effects on quality of life, universal literacy is important. The authors also detail how they have included artificial intelligence (AI) and augmented reality (AR) in their teaching of reading [16]. Mobile apps and devices equipped with these features facilitated interaction with written content and provided a more immersive learning environment. The findings point to the potential of technology to increase literacy opportunities for all.

According to the authors [17], they aim to demonstrate a comprehensive strategy to promote digital literacy among India's most marginalized rural population as part of the government's ambitious Digital India initiative. For low literacy learners in resource-poor environments with poor Internet bandwidth, lack of ICT facilities and inconsistent power, tackling multiple literacies at once poses a major challenge. The educational concept is an effective way to bring tablet-based digital literacy directly to communities, thus overcoming long-standing obstacles [18]. In order to improve both digital and life skills, it draws on a variety of actors,

including pre-existing civil society, schools, and government agencies, to deliver digital literacy and awareness. It demonstrates the benefits of a holistic approach to digital literacy as a tool to promote digital equity.

On the other hand, the authors [19] do a study to report on the design of specialized reading materials, such as books with altered fonts and simplified text. According to the findings, this type of content helped people with reading difficulties to read and understand what they read. The authors also discuss the benefits of cooperative and group learning for language acquisition. The researchers found that when people with reading difficulties worked together in small groups, they were able to encourage and support each other as they learned.

As more and more students with disabilities enroll in mainstream universities, the question of how best to accommodate them has become more pressing. The authors [20], define the support provided to these students remains a crucial task, despite the emphasis on inclusion and engagement in policy and practice. This collective case study used interviews and focus groups to gather information from 125 secondary school staff members from seven different schools about their experiences with students with disabilities [21]. Using the results of this research as a basis, a series of professional development initiatives were designed with the goal of improving the inclusion of older students with disabilities.

In this literature review, we analyzed the effectiveness and applicability of various techniques and a comprehensive synthesis of the most important findings. While it is true that the authors have expanded knowledge about intelligence, they have not offered any concrete plans on how to implement it using mobile applications.

### III. METHODOLOGY

Design Thinking is a method for identifying problems and proposing novel user-centered solutions. In this approach, priority is given to the requirements, desires, and feelings of the people for whom a product, service or experience is developed [22]. Product and service creation, as well as business innovation and complicated problem solving, are some of the areas in which this approach has been successful. Its human-centered approach and its ability to inspire innovation have led to its widespread adoption beyond the design world. The Design Thinking process is based on a series of five phases, the exact number of which will vary depending on the model or source used as shown in Fig. 1.

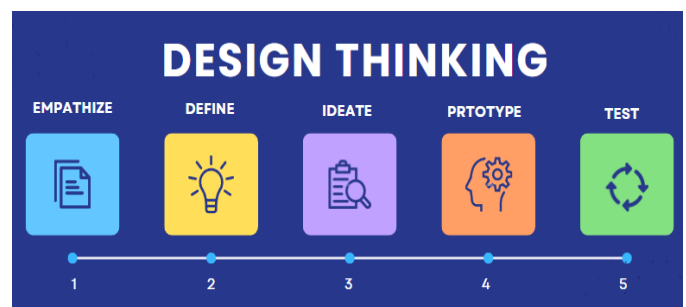


Fig. 1. Phases of design thinking.

A. Empathize

In this phase, the group investigates the target audience to learn more about their desires, needs, feelings and routines [23]. In doing so, it hopes to better understand the difficulties encountered by users. To do so, it can resort to interviews, participant observations and other ethnographic research methods. Table I shows the four questions posed for the interview with parents of people with reading disabilities.

TABLE I. INCLUSIVE LITERACY QUESTIONS

N°	Questions
1	Tell me about the learning and literacy experience of your child with a reading disability?
2	What kind of support or assistance has your child received to improve his/her reading skills?
3	Have you currently used any mobile apps or technology to support your child's literacy?
4	What specific content do you think would be most useful for your child in an inclusive literacy application?

B. Define

The "Define" phase of Design Thinking is the second step of the process and attempts to properly identify and characterize the problem or challenge that the design will address. A well-defined objective in this phase ensures that the rest of the design process moves in the right direction. Table II shows four questions posed for the survey to experts in special education using ICTs taking into account the interview report made in the previous step.

TABLE II. MOBILE APPLICATION QUESTIONS ON INCLUSIVE LITERACY

Dimensions	Questions
Support	Does the mobile application currently support you in developing your reading and comprehension skills more effectively?
Accessibility	Is the mobile application currently accessible for use by children with reading disabilities?
Monitoring	Are children frequently monitored by their teachers for academic progress in literacy in the proper use of the mobile application?
Motivation	Are special children motivated through appropriate strategies in the teaching and learning process by their teacher?

Table III shows the current status and prototype status of mobile applications for inclusive literacy for each of the dimensions such as support, accessibility, monitoring and motivation.

C. Ideas

The third step of Design Thinking, "Ideate" is about coming up with a wide range of original ideas and possible solutions to the problem or challenge identified in "Conceive" [24]. The goal at this point is to develop as many ideas as possible without worrying about their feasibility, thus encouraging diverse thinking, for the design of prototypes, about inclusive literacy. Table IV shows four consensual activities for the design of the mobile application for inclusive literacy in people with reading disabilities.

TABLE III. CURRENT AND PROPOSED SITUATION

Dimension	Current status	Proposed Situation
Support	We currently do not have a mobile support application.	To create a prototype of a mobile application to support learning.
Accessibility	Generally, there is no App available for its use.	Create a user-friendly App.
Monitoring	There is no frequent monitoring of your teacher.	There must be a monitoring and control plan
Motivation	No adequate motivation strategies	Strategies must be implemented in order to motivate

TABLE IV. CONSENSUAL ACTIVITIES FOR THE DESIGN OF THE MOBILE APPLICATION

N°	Consensual activities
Support	Design registration and Login to enter the inclusive literacy application for people with disabilities.
Accessibility	Conduct intuitive language and cognitive therapy inclusive literacy design for people with disabilities.
Monitoring	Design a mobile application of inclusive literacy educational games for people with disabilities.
	Design a mobile cognitive therapy application
Motivation	Design a mobile application for teaching reading for people with basic level disabilities
	Design a mobile application for teaching reading for people with intermediate level disabilities

D. Prototyping

The fourth step of Design Thinking is called "Prototyping" and its objective is to build rapid, low-cost prototypes of the solutions chosen in the "Ideate" phase [25]. Before settling on a final solution, ideas can be visualized and evaluated using these prototypes. At this point, the most important thing is to get feedback quickly and collect user feedback to help shape future iterations of the solutions.

Fig. 2 shows the Registration and Login of the inclusive literacy application. Users can access their own accounts and their own material by registering and logging into a mobile application. This function is essential to create a personalized and secure environment for all users.

Additionally, the registration and login process plays a pivotal role in enhancing user engagement and tracking individual progress within the app. By offering a personalized experience, users can easily pick up where they left off, track their achievements, and tailor their learning journey to meet their unique needs. This not only fosters a sense of ownership but also reinforces the commitment to fostering inclusive literacy, making the application a valuable tool for learners of all backgrounds and abilities.

Intuition-based cognitive and linguistic processing Designing for literacy inclusion requires a deep understanding of the strengths and weaknesses of people with cognitive or reading disabilities. Successful, accessible learning environments are the result of the combined efforts of educators, speech-language pathologists, and developers of

mobile apps and educational platforms. It helps people with disabilities improve their language and literacy skills by providing them with accessible information and tools. Images, visual elements and visual symbols are used extensively throughout the design to reinforce main points and improve readability. How to assemble vowels, combine by size and color (see Fig. 3).

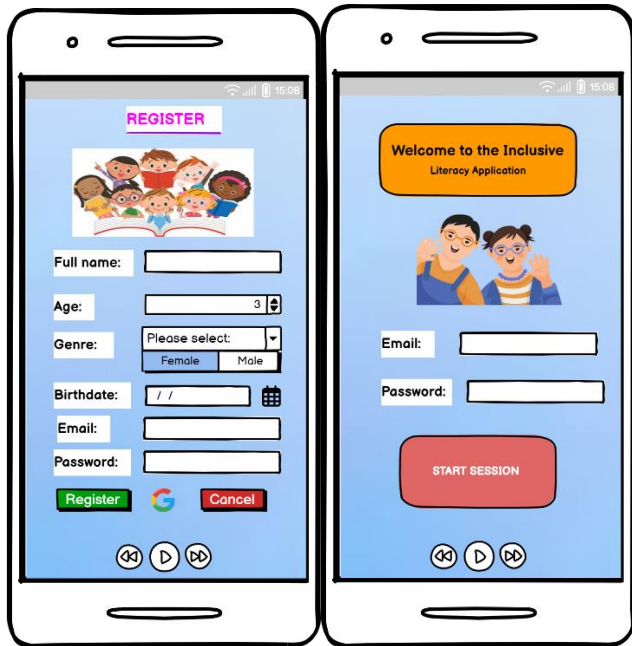


Fig. 2. Application registration and login.

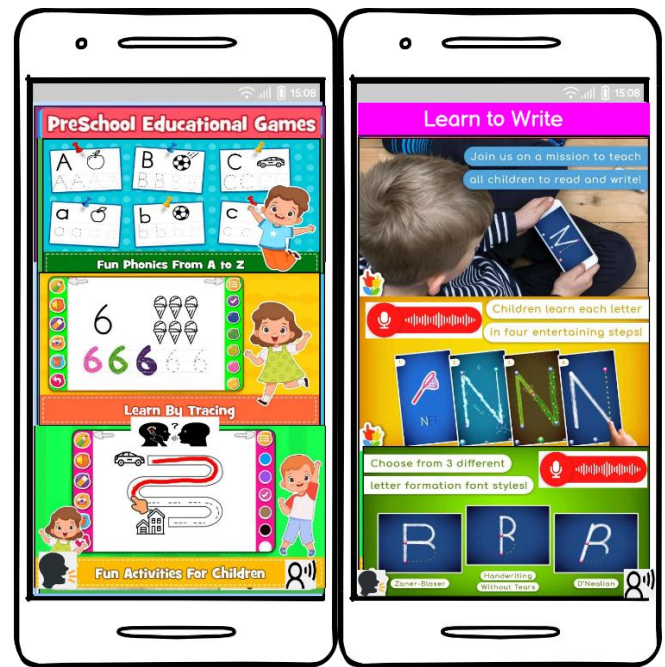


Fig. 4. Basic level educational games.

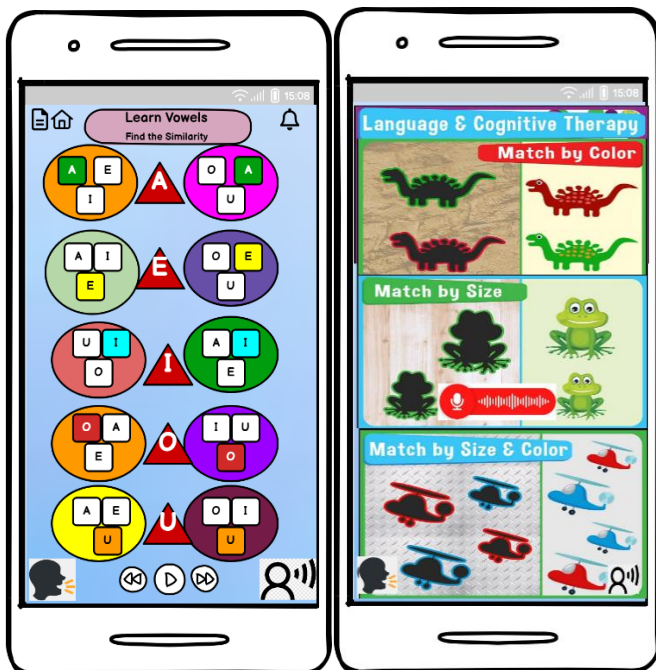


Fig. 3. Cognitive therapy.

Accessible, intelligible, and engaging learning experiences are the goal of the intuitive design of disability-inclusive literacy games. These video games take into account the specific needs of players to ensure that everyone has an equal opportunity to master the language, phonetic, letter recognition, and writing skills they offer. As shown in Fig. 4.

These inclusive educational games are modifiable to meet the needs of children with reading disabilities, giving them a voice in the learning process. These games provide a stimulating environment for children to learn to read and write using visual, auditory and tactile elements (see Fig. 5).

These inclusive educational games not only offer customization to cater to the specific requirements of children with reading disabilities but also play a crucial role in fostering inclusivity in the broader educational landscape. By accommodating diverse learning needs, they empower children with reading disabilities to actively participate in the learning process, promoting a sense of inclusion and equity. Additionally, these games serve as a dynamic and engaging platform for children to develop their literacy skills. Through a combination of visual, auditory, and tactile elements, they create a multisensory learning experience that appeals to various learning styles and strengths, further enhancing the accessibility and effectiveness of literacy instruction. Moreover, by embracing technology in education, these games align with the evolving digital era, preparing students with valuable digital literacy skills that are essential for their future success in a technology-driven world. In essence, these inclusive educational games not only bridge educational gaps but also nurture a more inclusive, adaptable, and tech-savvy generation of learners.

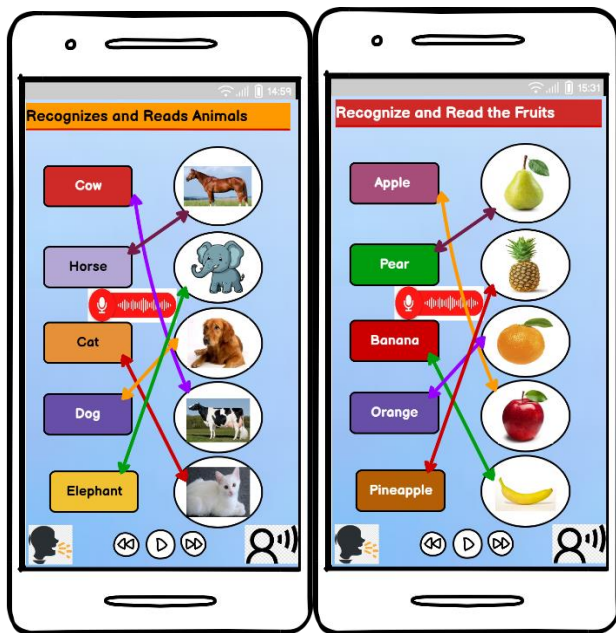


Fig. 5. Intermediate level educational games.

#### IV. RESULTS

##### A. About the Interview

Parents of people with reading disabilities were interviewed about inclusive literacy, especially in children, analyzing through Atlas TI 22 (See Fig. 6).

The utilization of Atlas.ti software in interviews holds significant importance as it empowers researchers to efficiently analyze and interpret qualitative data. This powerful tool facilitates the systematic organization of large volumes of textual, audio, or visual data, enabling researchers to uncover patterns, themes, and insights that might otherwise remain concealed. Its robust coding and data visualization capabilities not only streamline the research process but also enhance the rigor and credibility of qualitative studies, ultimately contributing to a deeper understanding of complex phenomena and more informed decision-making in various academic and professional domains.

- Experience:

It has been observed that all parents mention challenges in their children's learning experience, highlighting moments of frustration due to reading difficulties. However, a positive attitude toward each child's individual progress and their unique capacity to learn is also evident. This suggests that adaptability and a focus on individual achievements are essential aspects of the learning experience for these children.

- Support:

The analysis reveals that all parents have sought support and assistance both within the school environment and therapeutic settings. Speech and language therapy sessions emerge as a common intervention to enhance pronunciation

and auditory comprehension. Furthermore, collaboration with specialized tutors and the adaptation of learning environments are mentioned as effective strategies to address the unique needs of these children. This suggests that a combination of school-based and therapeutic approaches is crucial for the development of reading skills in children with reading disabilities.

- Use of Mobile Applications:

While all parents have experimented with mobile apps to support their children's literacy, there is a consistent search for a solution that perfectly aligns with the children's needs. Apps offering interactive exercises and voice narration have proven effective in maintaining interest and engagement. However, the lack of options perfectly tailored to the specific needs of the children underscores the importance of adaptability and customization in literacy apps.

- Content:

The analysis of parents' responses regarding useful content in an inclusive literacy app demonstrates a consensus on the significance of multimodality. The combination of text, images, and audio is essential to cater to diverse learning styles. Furthermore, interactive activities that reinforce vocabulary, sentence formation, and comprehension are considered valuable for enhancing reading skills. This highlights the need for tailored content addressing multiple aspects of literacy.

##### B. Expert Testing

Fifth, in the testing phase of Design Thinking, prototyped solutions are subjected to more rigorous testing with experts. In this phase, the solution is tested to gauge its interaction, usability, interface and quality before it is fully implemented. Expert validation of the design of the inclusive literacy prototypes was performed through evaluation with eight experts (E).

The solution undergoes testing to fine-tune its interaction, usability, interface, and quality before full implementation. The validation of the inclusive literacy prototype designs involved assessment by eight experts. The resulting average scores are as follows: For interaction, the mean was 81.25, indicating generally positive expert perceptions. Usability averaged 82.5, implying good overall usability. Interface garnered a high average of 91.25, reflecting strong approval of the design. The quality received an average of 82.5, denoting consistent good quality. Collectively, experts' evaluations suggest promising potential, yet addressing specific improvement areas noted by individual experts will further amplify their effectiveness and user experience (see Table V).

TABLE V. EXPERT VALIDATION

Criteria	E1	E2	E3	E4	E5	E6	E7	E8	Media
Interaction	80	90	80	90	70	80	80	80	81.25
Usability	70	80	90	80	80	90	80	90	82.50
Interface	80	90	90	100	100	90	90	90	91.25
Quality	80	80	80	90	80	80	80	90	82.50

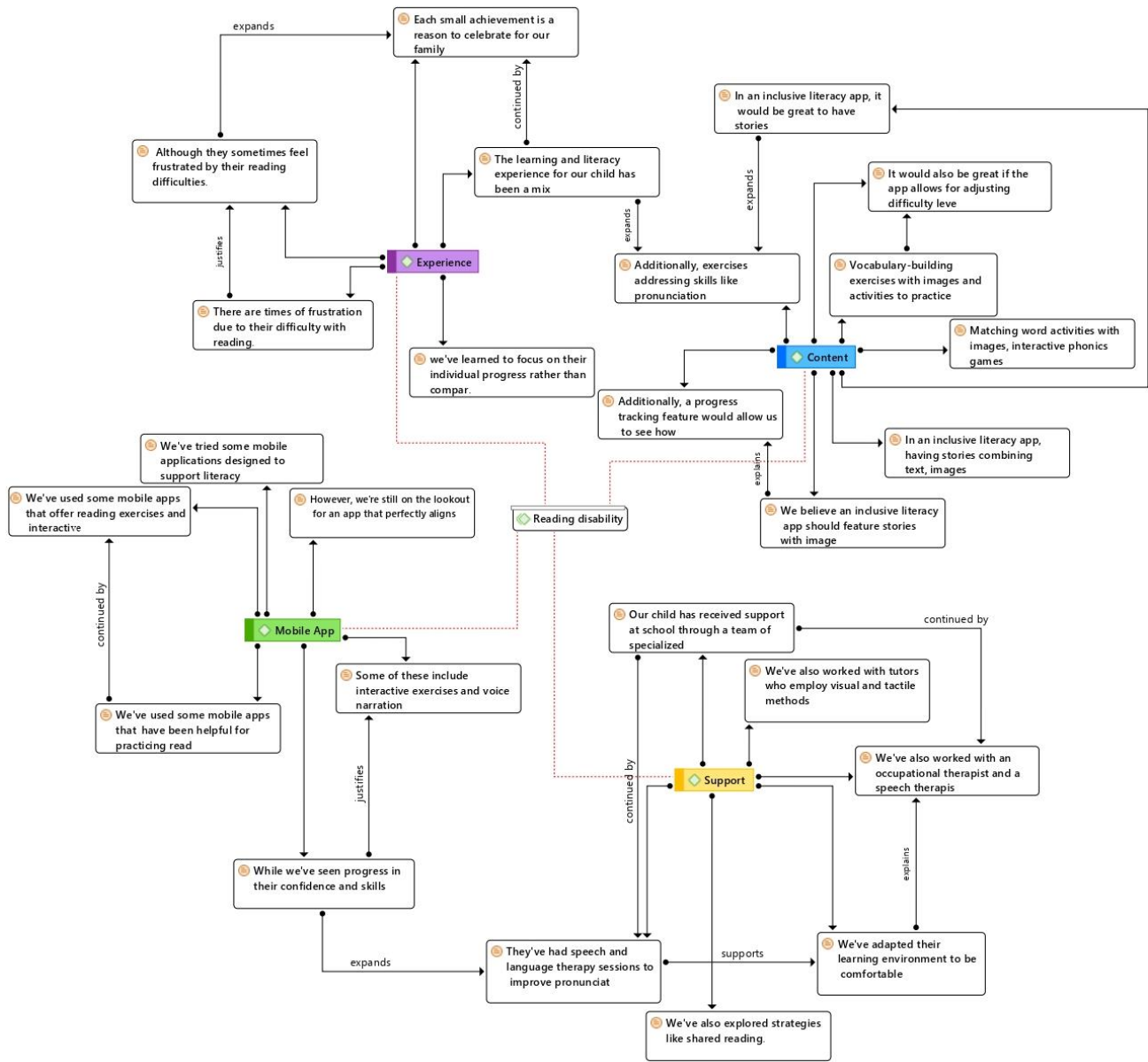


Fig. 6. Network analysis with the use of ATLAS ti22.

C. About the Survey

Table VI shows the four questions posed for the survey to parents about the use of the mobile application for inclusive literacy in people with reading disabilities.

TABLE VI. PARENT SURVEYS ON THE USE OF THE MOBILE APPLICATION

N°	Questions
1	Is the design of the mobile application user friendly?
2	Did the use of the mobile application serve as a complement for inclusive literacy?
3	Does the mobile application interact with people with reading disabilities?
4	Would you recommend the mobile application for inclusive literacy?

The analysis of the results from our prototype of a mobile application for inclusive literacy in individuals with reading

disabilities paints an encouraging picture. With an impressive 85%, the majority find the application design to be user-friendly, a vital aspect for creating a comfortable user experience. Furthermore, 75% confirm that the application effectively complements inclusive literacy efforts, a significant achievement in line with our goal. While 70% appreciate the application's interaction with individuals with reading disabilities, delving into the reasons behind the 30% who didn't share the same perception would be valuable. Lastly, an astounding 87% would gladly recommend the application, underscoring its valuable impact. While we're on the right track, it's crucial to address the feedback from those who didn't respond positively to continue refining and meeting their needs (see Fig. 7).

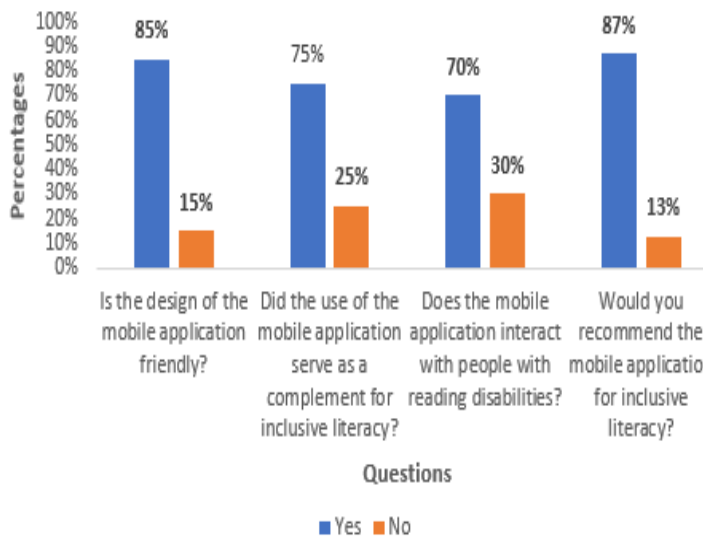


Fig. 7. Questions on the use of the mobile application.

## V. DISCUSSIONS

The authors [12] emphasize that reading comprehension and student engagement increased when individualized tactics such as guided reading and the use of pictograms were introduced. In the present research work, prototypes of didactic mobile applications specifically designed to enrich cognitive-behavioral therapy skills in people with reading disabilities were developed.

This study builds on previous research highlighting the importance of mobile applications and screen readers in promoting literacy among people with print disabilities [11]. Also, by adapting mobile applications for inclusive literacy to the needs of people with reading disabilities, the scope of these opportunities can be increased. Thus, in the present work, the contribution to the development of reading and writing skills, as well as to the full integration and participation of this population in society was realized.

This work was carried out based on interviews with parents and surveys of experts in special education with the use of ICTs. In the same way, the results were based on the inclusion of expert judgments and AtlasTI22 of the use of prototype mobile applications for inclusive literacy in people with reading disabilities. In contrast, the authors [15] did not conduct surveys and interviews, only studied people with reading disabilities to see how increasing their literacy levels affected their ability to relate to others and feel confident.

## VI. CONCLUSIONS AND FUTURE WORK

In conclusion, this study has investigated and developed the prototypes of mobile applications, with the aim of promoting inclusive literacy among people with reading difficulties. In order to evaluate the effectiveness and value of the proposed applications, interviews and surveys were conducted to collect the opinions of experts in the field to assess the effectiveness and usefulness of the proposed applications. The results obtained from expert judgment provide valuable insight into the feasibility and potential of these technological tools.

Analysis of the results of the prototype application designs yields an encouraging 85%. Similarly, 75% confirmed that the app effectively complements inclusive literacy efforts, a significant achievement in line with the objective, and 70% appreciated the app's interaction with people with reading disabilities. Finally, a staggering 87% would gladly recommend the app, underscoring its valuable impact. As for the methodology, design thinking was used as it is based on the human being approach, which addresses creativity, design and problem solving. A limitation of the research work is that it was not possible to contact directly the institutions of inclusive education to conduct interviews and make a qualitative analysis. As future work, it is recommended to implement the prototypes of mobile applications of inclusive education for people with reading disabilities complemented with augmented reality.

## REFERENCES

- [1] M. P. Campos, M. G. Retuerto, A. Delgado, and L. Andrade-Arenas, "Educational Platform to Improve Learning for Children with Autism," *International Journal of Engineering Pedagogy (iJEP)*, vol. 13, no. 2, pp. 20–35, Mar. 2023, doi: 10.3991/IJEP.V13I2.33969.
- [2] C. I. Martínez-Alcalá *et al.*, "Digital inclusion in older adults: A comparison between face-to-face and blended digital literacy workshops," *Frontiers in ICT*, vol. 5, no. AUG, p. 335246, Aug. 2018, doi: 10.3389/FICT.2018.00021/BIBTEX.
- [3] N. ChePa, N. Azzah Abu Bakar, L. Lim Sie-Yi, U. Utara Malaysia, and K. Malaysia, "Criteria and Guideline for Dyslexic Intervention Games," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 9, pp. 162–172, Dec. 2022, doi: 10.14569/IJACSA.2022.0130919.
- [4] W. Barber, "Inclusive and accessible physical education: rethinking ability and disability in pre-service teacher education," *Sport, Education and Society*, vol. 23, no. 6, pp. 520–532, Jul. 2016, doi: 10.1080/13573322.2016.1269004.
- [5] J. W. McKenna, M. Solis, F. Brigham, and R. Adamson, "The Responsible Inclusion of Students Receiving Special Education Services for Emotional Disturbance: Unraveling the Practice to Research Gap," *Behavior Modification*, vol. 43, no. 4, pp. 587–611, Mar. 2018, doi: 10.1177/0145445518762398.
- [6] N. I. Othman, N. A. M. Zin, and H. Mohamed, "Play-Centric Designing of a Serious Game Prototype for Low Vision Children," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 5, pp. 199–205, 2020, doi: 10.14569/IJACSA.2020.0110528.
- [7] F. Kiuppis, "Inclusion in sport: disability and participation," *Sport in society*, vol. 21, no. 1, pp. 4–21, Jan. 2016, doi: 10.1080/17430437.2016.1225882.
- [8] S. A. Boyle, D. McNaughton, and S. E. Chapin, "Effects of Shared Reading on the Early Language and Literacy Skills of Children With Autism Spectrum Disorders: A Systematic Review," *Focus on Autism and Other Developmental Disabilities*, vol. 34, no. 4, pp. 205–214, May 2019, doi: 10.1177/1088357619838276.
- [9] S. G. Wood, J. H. Moxley, E. L. Tighe, and R. K. Wagner, "Does Use of Text-to-Speech and Related Read-Aloud Tools Improve Reading Comprehension for Students With Reading Disabilities? A Meta-Analysis," *Journal of learning disabilities*, vol. 51, no. 1, pp. 73–84, Jan. 2017, doi: 10.1177/0022219416688170.
- [10] D. Chadwick *et al.*, "Digital inclusion and participation of people with intellectual disabilities during COVID-19: A rapid review and international bricolage," *Journal of Policy and Practice in Intellectual Disabilities*, vol. 19, no. 3, pp. 242–256, Sep. 2022, doi: 10.1111/JPPI.12410.
- [11] L. Stinken-Rösner *et al.*, "Thinking Inclusive Science Education from two Perspectives: inclusive Pedagogy and Science Education RISTAL 3 / 2020 Research in Subject-matter," *RISTAL. Research in Subject-matter Teaching and Learning*, vol. 3, pp. 30–45, 2020, doi: 10.23770/r1831.

- [12] K. Ishaq, N. A. M. Zin, F. Rosdi, A. Abid, and Q. Ali, "Usefulness of Mobile Assisted Language Learning in Primary Education," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 1, pp. 384–395, Spring 2020, doi: 10.14569/IJACSA.2020.0110148.
- [13] A. Jalil, T. Tohara, S. M. Shuhidan, F. Diana, S. Bahry, and M. Norazmi Bin Nordin, "Exploring Digital Literacy Strategies for Students with Special Educational Needs in the Digital Age," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 9, pp. 3345–3358, Apr. 2021, doi: 10.17762/TURCOMAT.V12I9.5741.
- [14] A. Sánchez-Morales, J. A. Durand-Rivera, and C. L. Martínez-González, "Usability Evaluation of a Tangible User Interface and Serious Game for Identification of Cognitive Deficiencies in Preschool Children," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 6, pp. 486–493, 2020, doi: 10.14569/IJACSA.2020.0110661.
- [15] B. Xie, N. Charness, K. Fingerma, J. Kaye, M. T. Kim, and A. Khurshid, "When Going Digital Becomes a Necessity: Ensuring Older Adults' Needs for Information, Services, and Social Inclusion During COVID-19," *Older Adults and COVID-19*, vol. 32, no. 4–5, pp. 460–470, Jul. 2020, doi: 10.1080/08959420.2020.1771237.
- [16] I. Koomson, R. A. Villano, and D. Hadley, "Intensifying financial inclusion through the provision of financial literacy training: a gendered perspective," *Applied Economics*, vol. 52, no. 4, pp. 375–387, Jan. 2019, doi: 10.1080/00036846.2019.1645943.
- [17] S. Druga, S. T. Vu, E. Likhith, and T. Qiu, "Inclusive AI literacy for kids around the world," ACM International Conference Proceeding Series, pp. 104–111, Mar. 2019, doi: 10.1145/3311890.3311904.
- [18] K. de Bruin, "The impact of inclusive education reforms on students with disability: an international comparison," *International Journal of inclusive education*, vol. 23, no. 7–8, pp. 811–826, Aug. 2019, doi: 10.1080/13603116.2019.1623327.
- [19] J. J. Murray, K. Snoddon, M. De Meulder, and K. Underwood, "Intersectional inclusion for deaf learners: moving beyond General Comment no. 4 on Article 24 of the United Nations Convention on the Rights of Persons with Disabilities," *International Journal of Inclusive Education*, vol. 24, no. 7, pp. 691–705, Jun. 2018, doi: 10.1080/13603116.2018.1482013.
- [20] D. Maciver, C. Hunter, A. Adamson, Z. Grayson, K. Forsyth, and I. McLeod, "Supporting successful inclusive practices for learners with disabilities in high schools: a multisite, mixed method collective case study," *Disability and rehabilitation*, vol. 40, no. 14, pp. 1708–1717, Jul. 2017, doi: 10.1080/09638288.2017.1306586.
- [21] A. Kart and M. Kart, "Academic and Social Effects of Inclusion on Students without Disabilities: A Review of the Literature," *Education Sciences 2021, Vol. 11, Page 16*, vol. 11, no. 1, p. 16, Jan. 2021, doi: 10.3390/EDUCSCI11010016.
- [22] K. D. Elsbach and I. Stigliani, "Design Thinking and Organizational Culture: A Review and Framework for Future Research," *Journal of Management*, vol. 44, no. 6, pp. 2274–2306, Jan. 2018, doi: 10.1177/0149206317744252.
- [23] P. Micheli, S. J. S. Wilner, S. H. Bhatti, M. Mura, and M. B. Beverland, "Doing Design Thinking: Conceptual Review, Synthesis, and Research Agenda," *Journal of Product Innovation Management*, vol. 36, no. 2, pp. 124–148, Mar. 2019, doi: 10.1111/JPIM.12466.
- [24] M. Altman, T. T. K. Huang, and J. Y. Breland, "Peer Reviewed: Design Thinking in Health Care," *Prev Chronic Dis*, vol. 15, no. 9, p. 180128, Sep. 2018, doi: 10.5888/PCD15.180128.
- [25] J. P. S. Bartra, J. F. H. Puja, M. G. Retuerto, and L. Andrade-Arenas, "Prototype of Mobile Application Oriented to the Educational Help for Blind People in Peru," *International Journal of Interactive Mobile Technologies (IJIM)*, vol. 16, no. 17, pp. 130–147, Sep. 2022, doi: 10.3991/IJIM.V16I17.32075.

# LAD-YOLO: A Lightweight YOLOv5 Network for Surface Defect Detection on Aluminum Profiles

Dongxue Zhao, Shenbo Liu, Yuanhang Chen, Da Chen, Zhelun Hu, Lijun Tang

School of Physics and Electronic Science, Changsha University of Science & Technology, Changsha 410114, China

**Abstract**—In this paper, we leverage the advantages of YOLOv5 in target detection to propose a highly accurate and lightweight network, called LAD-YOLO, for surface defect detection on aluminum profiles. The LAD-YOLO addresses the issues of computational complexity, low precision, and a large number of model parameters encountered in YOLOv5 when applied to aluminum profiles defect detection. LAD-YOLO reduces the model parameters and computation while also decreasing the model size by utilizing the ShuffleNetV2 module and depthwise separable convolution in the backbone and neck networks, respectively. Meanwhile, a lightweight structure called "Ghost\_SPPFCSPC\_group", which combines Cross Stage Partial Network Connection Operation, Ghost Convolution, Group Convolution and Spatial Pyramid Pooling-Fast structure, is designed. This structure is incorporated into the backbone along with the Convolutional Block Attention Module (CBAM) to achieve lightweight. Simultaneously, it enhances the model's ability to extract features of weak and small targets and improves its capability to learn information at different scales. The experimental results show that the mean Average Precision (mAP) of LAD-YOLO on aluminum profiles defect datasets reaches 96.9%, model size is 6.64MB, and Giga Floating Point Operations (GFLOPs) is 5.5. Compared with YOLOv5, YOLOv5s-MobileNetv3, and other networks, LAD-YOLO proposed in this paper has higher accuracy, fewer parameters, and lower floating-point computation.

**Keywords**—YOLOv5; ShuffleNetv2; lightweight and fast spatial pyramid pooling structure; convolutional block attention module; aluminum profiles surface defect detection

## I. INTRODUCTION

Aluminum profiles are one of the important raw materials for the manufacturing industry, widely used in industry, construction, medicine, and other industries. However, due to its complex production process and more transportation links, aluminum profiles are prone to surface defects such as scratch, dirt, pinhole, and wrinkle. These defects will directly affect the quality of aluminum profiles and even lead to distortion and deformation of aluminum profiles, which is more obvious for high-end aluminum profiles. Therefore, it is of great significance to improve the detection efficiency and accuracy of aluminum profiles surface defects to ensure the production and application of aluminum profiles.

At present, most enterprises still use traditional manual detection methods to detect defects on the surface of aluminum profiles. However, this manual inspection method is slow, subject to the influence of subjective consciousness, not only low efficiency but also poor stability, prone to misdetection, and leakage detection. Ultrasonic flaw detection, eddy current

flaw detection, and other traditional non-destructive testing are also used for the detection of surface defects in aluminum profiles, but due to its slow detection speed, high cost, complex equipment operation, etc., which limits its popularity in practical applications. In 2014, Girshick et al. proposed a Regional Convolutional Neural Network (R-CNN), which broke the deadlock of slow progress in the field of target detection [1], and subsequently gave birth to Fast R-CNN [2], Faster R-CNN [3], Mask R-CNN [4], Single Shot MultiBox Detector (SSD) [5], You Only Look Once (YOLO) series [6-11], and other generalized deep learning-based target detection algorithms. As a result, deep learning-based surface defect detection is starting to develop rapidly.

For metal surface defect detection, references [12-14] combined neural networks with traditional detection algorithms to realize the detection and classification of surface defects of aluminum and other metal materials. Duan et al. [15] built a dual-stream Convolutional Neural Network (CNN) for the detection of aluminum profiles image features and gradient features, effectively realizing the classification of defect-free and multi-type defect samples. Cheng et al. [16] proposed a network DEA-RetinaNet with differential channel attention and adaptive spatial feature fusion for steel surface defect detection. The mean Average Precision (mAP) of the network on the steel surface defect dataset (NEU-DET) was 78.25%. The detection accuracy of the above methods is lower than 85%, which cannot meet the requirements of practical industrial applications.

Zeng [17] et al. proposed a data augmentation method and a migration learning technique for solving defective parts detection in steel plates. References [18-20] used Faster R-CNN to detect metal surface defects such as steel and railroad fasteners with an accuracy of more than 95%, but the detection speed is slow. Chen [21] et al. applied Convolutional Neural Networks (DCNNs) to the defect detection of fasteners and carried out experiments on high-speed railroad scenarios. ZHAO et al. [22] innovated based on YOLOv4 architecture to improve the detection accuracy of surface defects of metal materials. Wang et al. [23] proposed a structure called PE-Neck, which replaces the Neck part of the YOLOv5 network structure with a combination of scaled convolutional kernels and efficient channel attention to enhance the model's ability to extract and localize defects at different scales. However, the accuracy is only 87.4% and the strategy for generating candidate regions suffers from many flaws. Although the above methods enhance the detection accuracy by improved means, it is unable to realize the real-time detection of surface defects on



industrial aluminum profiles due to their complex network structure and large computation, and slow detection speed.

Conventional CNN inference is computationally intensive and difficult to apply in resource-constrained scenarios such as mobile and Internet of Things (IoT). Starting from SqueezeNet [24] and MobileNetV1 [25], the design of CNNs has begun to focus on efficiency in resource-constrained scenarios. The more mature lightweight networks include the MobileNet series [26-27], ShuffleNet series [28-29], GhostNet [30], etc. Li [31] et al. proposed a YOLOv3-Lite detection method, which combines a deep convolutional neural network and a feature pyramid in YOLOv3, to improve the defect detection accuracy. Xiao [32] et al. added a residual network structure to YOLOv3-Tiny, which was applied to detect obstacles in a mine, with improved accuracy compared to the original YOLOv3-Tiny, but with decreased speed. Zhang [33] et al. proposed a multi-model rail surface defect detection system based on a convolutional neural network (MRSDI-CNN). The system network uses SDD combined with YOLOv3 to improve the system's accuracy. Wang et al [34] proposed a lightweight YOLO-ACG detection algorithm that balances accuracy and speed while improving the defect detection classification error and leakage rate. Ma [35] et al improved the YOLOv4 network by replacing the backbone network with a lightweight Ghost module. At the same time, a joint attention mechanism is added to the stacked Ghost modules to ensure accuracy, so that the network is compressed and lightweight is achieved while achieving an accuracy of 94.68%. These methods have improved in detection accuracy and detection speed, but the model size is still large and memory consumption is high, which is not conducive to real-time detection on mobile, especially in devices with tight computing resources.

The YOLOv5 algorithm is an end-to-end target detection algorithm known for its fast detection speed and high accuracy. It has found wide application in the field of surface defect detection. However, the large number of parameters in the YOLOv5 model can hinder improvements in detection speed. Additionally, its backbone layer, consisting of CSPDarknet53, faces challenges in effectively extracting features of small

targets. In this paper, a lightweight aluminum profiles surface defect detection network is designed to solve this problem, which significantly improves the accuracy and detection speed. The algorithm is evaluated for its performance on aluminum profiles surface defect dataset and compared with other algorithms. The experimental results show that the LAD-YOLO proposed in this study can accurately identify aluminum profiles surface defects with excellent detection speed.

The Section I is the research purpose and significance of aluminum profiles surface defects detection and the current status of domestic and international research on target detection algorithms for metal surface defects detection. The Section II is the research on the improvement method of lightweight aluminum profiles surface defect detection model based on YOLOv5. The Section III is the experimental results and analysis of the model application. The Section IV summarizes the research in this paper and the outlook for future research.

## II. METHODOLOGY

LAD-YOLO follows the network structure of YOLOv5, which consists of four main parts: Input Layer, Backbone, Neck, and Head. The overall structure of LAD-YOLO is depicted in Fig. 1. The Input Layer takes a  $640 \times 640 \times 3$  aluminum profiles defect image as input. The Backbone network contains six ShuffleNetv2 modules, three Convolutional Block Attention Modules (CBAM), and the Spatial Pyramid Pooling Cross Stage Partial Concat structure based on ghost convolution and group convolution (Ghost\_SPPFCSPC\_group) for extracting surface defect features of aluminum profiles. In the neck network, depthwise separable convolution is used to extract depth features of aluminum profiles surface defects, reducing computational overhead. Simultaneously, the feature image size is doubled by using the nearest neighbor interpolation upsampling method, and the feature maps with the same size in the aluminum profiles surface defect map are connected. In the prediction layer, three different sizes of detection heads are generated to detect the aluminum profiles' surface defect image.

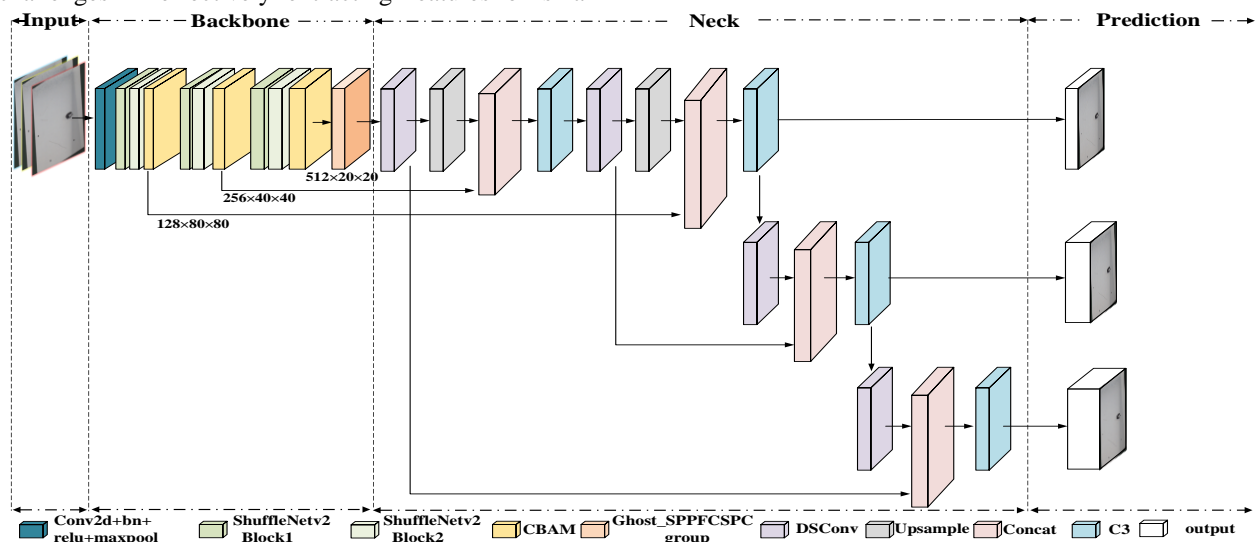


Fig. 1. LAD-YOLO network structure.

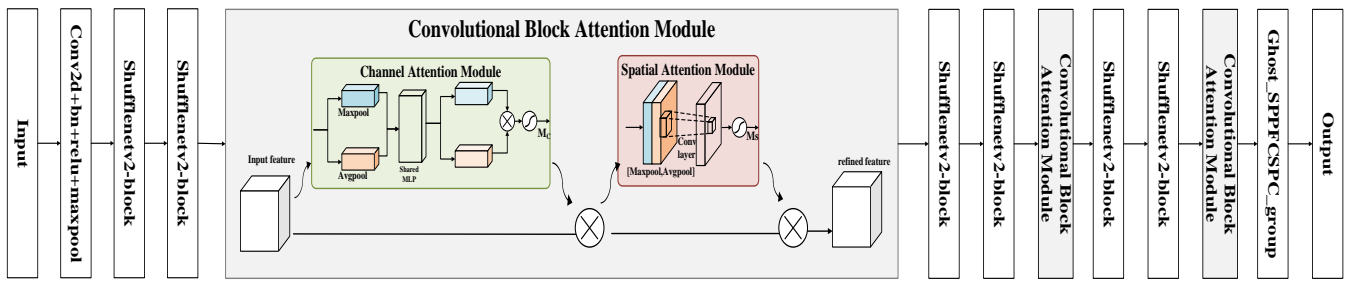


Fig. 2. CBAM and LAD-YOLO backbone layer network structure.

A. Backbone Structure

The Backbone Structure shown in Fig. 2 consists of three parts: ShuffleNetv2, Convolutional Block Attention Module (CBAM), and the Spatial Pyramid Pooling-Fast Cross Stage Partial Concat structure combining Ghost Convolution and Group Convolution (Ghost\_SPPFCSPC\_group). ShuffleNetv2 uses Channel Split, 1\*1 convolution, depthwise separable convolution, and mixing and washing of channels to accomplish the detection of input aluminum profiles defect information, which reduces the memory access time, reduces the number of model parameters, and improves the detection speed. The adoption of the ShuffleNetv2 structure drastically reduces the number of parameters of the model, but also brings a certain loss of accuracy. To compensate for the loss of accuracy, CBAM is used to embed into the backbone, as shown in Fig. 2.

CBAM mainly consists of two key modules: the Channel Attention Module and the Spatial Attention Module. The Channel Attention Module captures the importance of each feature channel by calculating global statistics and applies attention weights to each channel. The Spatial Attention Module highlights important spatial regions in the feature map by computing global statistics and applying attention weights to different spatial positions. Combining the above two modules, CBAM enables the network to adaptively focus on significant channels and spatial areas, improving feature representation for aluminum profiles surface defect detection tasks.

B. Ghost\_SPPFCSPC\_Group Structure

The Spatial Pyramid Pooling (SPP) structure can effectively capture target features at different scales by stacking pyramid layers of different sizes together, improving the model's detection ability for targets of different sizes. The Spatial Pyramid Pooling-Fast (SPPF) structure is a faster structure proposed based on the SPP structure. The Cross Stage Partial structure consists of two parts, the convolution, and the complex structure, in parallel to increase the speed of the network.

In 2023, wang et al. [11] first proposed the Spatial Pyramid Pooling Cross Stage Partial Concat (SPPCSPC) structure in YOLOv7, as shown in Fig. 3, which uses the SPP and CSP modules for better handling of multi-scale targets.

The use of the SPPCSPC structure can effectively improve the model detection accuracy, but it will increase the amount of computation and the number of model parameters. To improve the speed, this paper replaces the SPP in the SPPCSPC

structure with the SPPF structure to obtain the Spatial Pyramid Pooling-Fast Cross Stage Partial Concat (SPPFCSPC) structure. To reduce the amount of computation and parameters, Ghost convolution and group convolution are used to replace standard convolution in the SPPFCSPC structure.

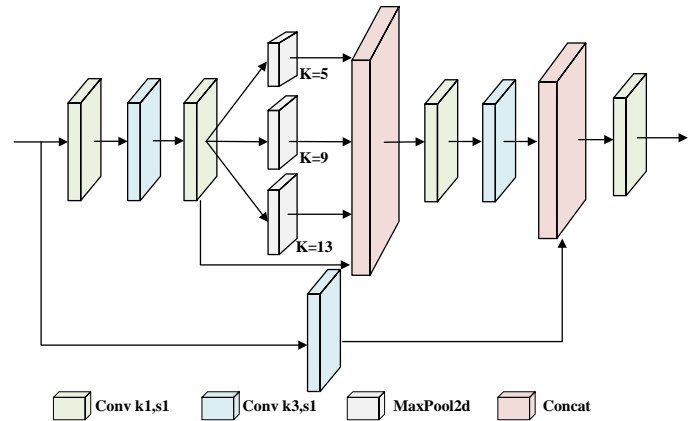


Fig. 3. Spatial pyramid pooling cross stage partial concat structure.

The standard convolution, Ghost convolution, and group convolution are compared and analyzed below. Fig. 4 shows the operation process of standard convolution, Eq. (1) is the standard convolution parameters, where  $c$  is the number of input channels,  $n$  is the number of output channels, and the size of the convolution kernel is  $k*k$ .

$$P_{std} = c \cdot n \cdot k \cdot k \tag{1}$$

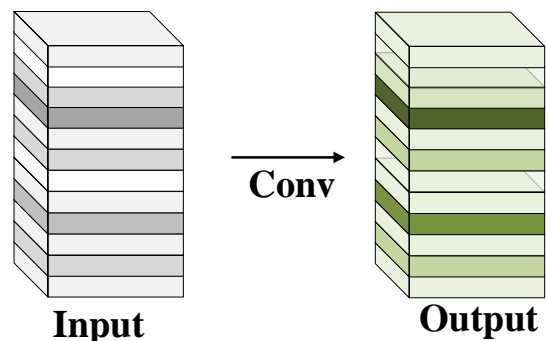


Fig. 4. Standard convolution operation.

Fig. 5 shows the group convolution operation process, which divides the input channels and output channels into the same number of groups, and then allows the input channels and output channels in the same group number to be fully

connected. Eq. (2) is the parameters of the group convolution. Where  $g$  is the number of groups divided into output channels.

$$P_{Group} = \frac{c}{g} \cdot \frac{n}{g} \cdot g \cdot k \cdot k \quad (2)$$

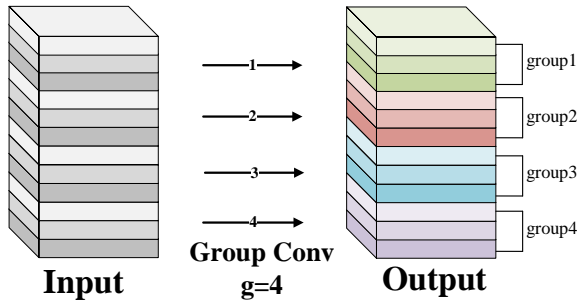


Fig. 5. Group convolution operation.

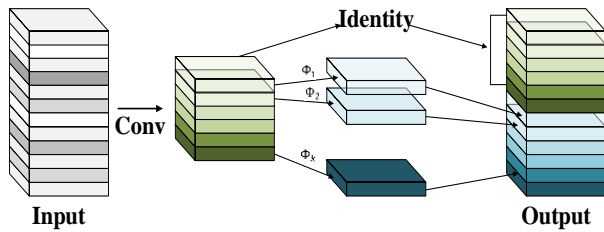


Fig. 6. Ghost convolution operation.

Fig. 6 shows the Ghost convolution operation process. First, a standard convolution operation with  $m$  ( $m \leq n$ ) output channels (where  $n$  is the number of final output channels) is performed on the feature map with input channel  $c$  to obtain a feature map with  $m$  channels. Second, a new feature map is obtained by  $s-1$

linear operations. Finally, the two feature maps are connected to obtain an output feature map with  $n$  channels ( $n=m*s$ ). Eq. (3) is the parameters of Ghost convolution. Where the convolution kernel size of linear operations in Ghost Module is  $d \times d$ .

$$P_{Ghost} = \frac{n}{s} \cdot c \cdot k \cdot k + (s-1) \cdot \frac{n}{s} \cdot d \cdot d \quad (3)$$

By comparing the above three convolution parameters, Eq. (4) shows the parameters compression ratio ( $R_1$ ) for standard and Ghost convolution, and Eq. (5) shows the parameters compression ratio ( $R_2$ ) for standard and group convolution. Where let  $k=d$ .

$$R_1 = \frac{n \cdot c \cdot k \cdot k}{\frac{n}{s} \cdot c \cdot k \cdot k + (s-1) \cdot \frac{n}{s} \cdot d \cdot d} \approx \frac{s \cdot c}{s+c-1} \approx s \quad (4)$$

$$R_2 = \frac{n \cdot c \cdot k \cdot k}{\frac{c \cdot n}{g} \cdot g \cdot k \cdot k} = g \quad (5)$$

From the above computation results, it is shown that the number of parameters of standard convolution is  $g$  times more than that of group convolution and  $s$  times more than that of Ghost convolution, so the number of parameters can be drastically reduced by choosing group convolution and Ghost convolution compared to standard convolution.

Therefore, in this paper, a lightweight SPPFCSPC structure (Ghost\_SPPFCSPC\_group) is designed by combining SPPFCSPC, Ghost convolution, and group convolution, as shown in Fig. 7. The structure utilizes the smaller number of parameters of Ghost convolution and group convolution to achieve lightweight. The Ghost\_SPPFCSPC\_group structure uses a smaller computational cost to enable the fusion of multi-scale features and improve feature representation.

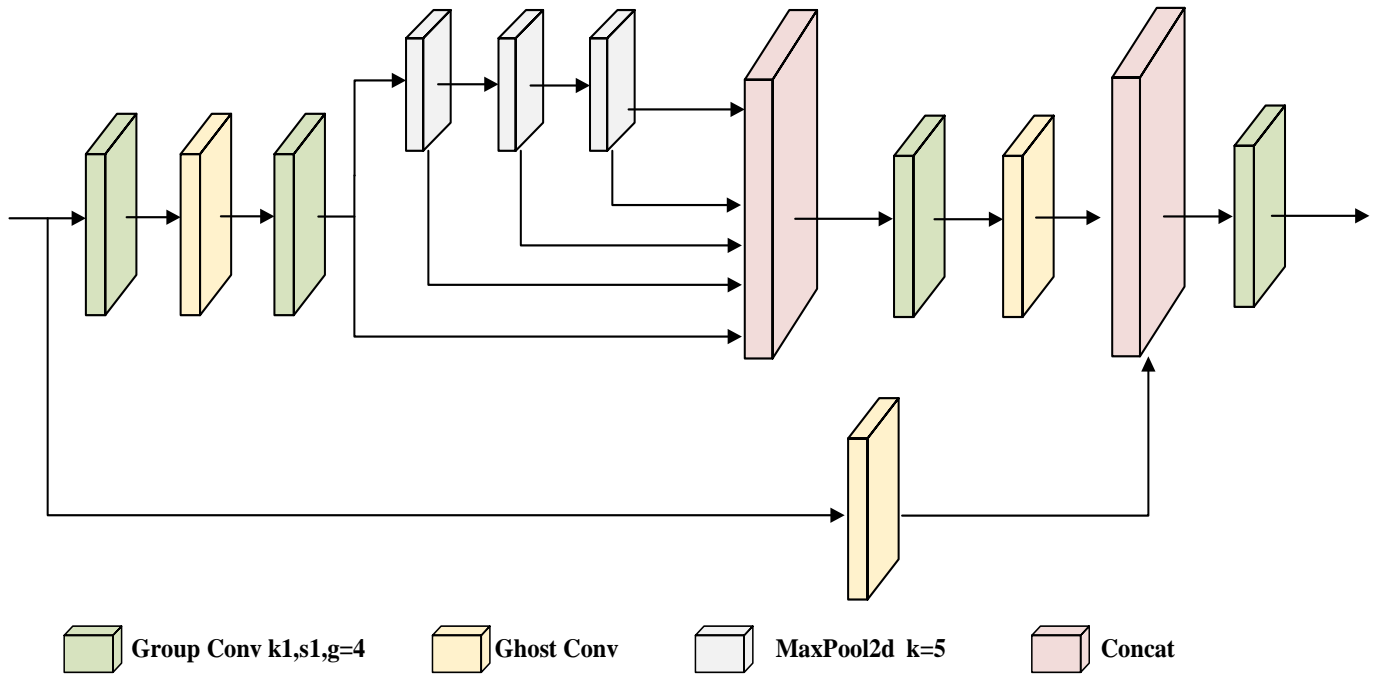


Fig. 7. Ghost\_SPPFCSPC\_group structure.

### C. Depthwise Separable Convolution

Depthwise Separable Convolution (DSCConv) contains two parts, Depthwise Convolution and Pointwise Convolution. As shown in Fig. 8, Depthwise Convolution computes the convolution of each channel separately to extract the features of each channel; Pointwise Convolution computes the feature map generated by Depthwise Convolution and adopts a convolution kernel with the size of  $1 \times 1 \times M$  convolution kernel, weighted combination in the depth direction, to realize the fusion of features between the channels, to generate a new feature map.

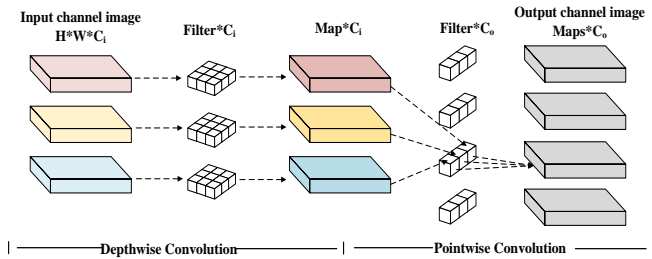


Fig. 8. Depthwise separable convolution module structure.

Eq. (6) is the FLOPs for standard convolution. Eq. (7) is the FLOPs Operations for Depthwise Separable Convolution. Eq. (8) is the ratio  $R$  of Depthwise Separable Convolution to the standard convolutional computation. The input feature map size is  $H \times W \times C_i$ , the output feature map size is  $H \times W \times C_o$ ,  $H$ ,  $W$ ,  $C_i$  and  $C_o$  denote the height, width, number of input channels, and number of output channels of the feature map respectively, and the size of the convolution kernel in the standard convolution is  $K_1 \times K_2$ .

$$C_{std} = K_1 \cdot K_2 \cdot H \cdot W \cdot C_i \quad (6)$$

$$\begin{aligned} C_{separable} &= C_{depthwise} + C_{pointwise} \\ &= K_1 \cdot K_2 \cdot H \cdot W \cdot C_i + H \cdot W \cdot C_i \cdot C_o \end{aligned} \quad (7)$$

$$\begin{aligned} R &= \frac{C_{separable}}{C_{std}} = \frac{K_1 \cdot K_2 \cdot H \cdot W \cdot C_i + H \cdot W \cdot C_i \cdot C_o}{K_1 \cdot K_2 \cdot H \cdot W \cdot C_i \cdot C_o} \\ &= \frac{1}{C_o} + \frac{1}{K_1 \cdot K_2} \end{aligned} \quad (8)$$

From Eq. (8), it can be seen that the floating-point computation of Depthwise Separable Convolution is only  $\frac{1}{C_o} + \frac{1}{K_1 \cdot K_2}$  of the standard convolution. Assuming the convolution kernel size of  $3 \times 3$  for Depthwise Convolution, the computation of the standard convolution is about eight to nine times that of Depthwise Separable Convolution. Replacing the standard convolution with the Depthwise Separable Convolution reduces the floating-point computation.

## III. RESULTS AND DISCUSSION

### A. Datasets Introduction

In this paper, for aluminum profiles defect detection, the Hikvision high-definition industrial camera model MV-CS050-10GC-PRO is used to collect the sample images of aluminum profiles and make the aluminum profiles defect datasets, and some of the data in the datasets are shown in Fig. 9.

The labeling categories are pinhole, scratch, dirt, and wrinkle. Since the original image samples are too few, panning, rotating, changing brightness, shearing, mirroring, and other means of expanding the datasets are chosen to expand the original aluminum profiles surface defect datasets. After the expansion, the total number of defect images is 5013, including 6325 pinholes, 3042 dirt, 5863 scratches, and 2415 wrinkles. The datasets are categorized into 60% training set, 20% validation set, and 20% test set containing 3008, 1002, and 1003 images, respectively.



Fig. 9. Part of the datasets.

### B. Evaluation Metrics

Precision (P), Recall (R), Average Precision (AP), and Mean Average Precision (mAP) are used as the evaluation metrics of detection effectiveness. The mAP is the average value of AP for all defect categories, which is used as a comprehensive index for evaluating precision. The higher the values of AP and mAP, the better the algorithm is for detecting the target defects. P, R, AP, and mAP are calculated as follows: (9), (10), (11), (12).  $TP$  is the number of defects in the positive samples that were detected as correct,  $FP$  is the number of defects in the negative samples that were incorrectly detected as correct,  $FN$  is the number of defects in the positive samples that were not detected, and  $m$  is the number of defect categories. In addition, Floating Point Operations (FLOPs), parameters, and Model Size are used to evaluate the lightness of the model.

$$P = \frac{TP}{TP+FP} \quad (9)$$

$$R = \frac{TP}{TP+FN} \quad (10)$$

$$AP_i = \int_0^1 PRdr \quad (11)$$

$$mAP = \frac{1}{m} \sum_{i=1}^m AP_i \quad (12)$$

### C. Experimental Process

The GPU used for defect detection training and testing is NVIDIA TITAN RTX, and the specific configuration of the experimental platform is shown in Table I. During the training

experiments, the optimizer chooses the stochastic gradient descent with momentum (SGD) with a momentum factor of 0.937. The weight attenuation coefficient is set to  $5 \times 10^{-4}$ . The learning rate is initially set to  $10^{-3}$ , while the Cosine Annealing is used to reduce the learning rate to  $10^{-5}$ . The batch size is set to 64, and the epochs are set to 500.

TABLE I. THE SPECIFIC CONFIGURATIONS OF THE EXPERIMENTAL PLATFORM

Name	Version
CPU	Intel(R) Xeon(R) Gold 5218R CPU @ 2.10GHz
Memory Bank	32GB
GPU	NVIDIA TITAN RTX
GPU Memory	24GB
Operating System	Windows10
Software environment	CUDA11.6
Python Version	Python 3.8
Deep learning framework	PyTorch 1.12

#### D. Comparative Experiment and Analysis

Comparison experiments are conducted by LAD-YOLO with SSD, YOLOv3, YOLOv3-tiny, YOLOv4-tiny, YOLOv5, and YOLOv5s-MobileNetv3, and the results are shown in Table II. As can be seen in Table II, LAD-YOLO achieved 96.9% mAP, 97.4% Precision, and 95.7% Recall on the aluminum profiles surface defects dataset, and the model size is 6.64MB. Compared with the YOLOv5s algorithm, mAP increases by 2.8% and model size decreases by 58%.

TABLE II. COMPARATIVE RESULTS OF EVALUATION METRICS FOR DIFFERENT METHODOLOGIES

Methods	Precision (%)	Recall (%)	mAP (%)	Model Size (MB)
SSD	68.4	70.2	70.6	90.13
YOLOv3	91.9	87.7	91.2	120.67
YOLOv3-tiny	87.6	84.7	86.9	33.79
YOLOv4-tiny	90.9	89.6	90.8	23.03
YOLOv5s	95.3	94.0	94.1	14.07
YOLOv5s-MobileNetv3	89.9	88.4	89.7	7.28
LAD-YOLO (OURS)	97.0	95.7	96.9	6.64

TABLE III. RESULTS OF THE ABLATION EXPERIMENT

ShuffleNetV2	CBAM	SPPFCSPC	Ghost_SPPFCSPC_group	DSCConv	mAP	Parameters (10 <sup>6</sup> )	Model Size (MB)	GFLOPs
--	--	--	--	--	94.1%	7.03	14.08	15.8
√					92.2%	3.23	6.66	5.8
√	√				95.8%	3.25	6.69	5.9
√	√	√			97.3%	9.77	19.46	11.1
√	√		√		97.0%	3.85	7.91	6.3
√	√		√	√	96.9%	3.19	6.64	5.5

Compared with YOLOv5s-MobileNetv3, the precision is improved by 7.2% and the model size is reduced by 8.8%. The experimental results show that the LAD-YOLO network improves the precision and recall rate of defect detection, and reduces the model parameters and size.

#### E. Ablation Experiment

To further verify the role of each improvement in enhancing the performance of the algorithm, ablation experiments are conducted. The results are shown in Table III.

From Table III, the mAP of the baseline model YOLOv5s is 94.1%, the model size is 14,08MB, the number of parameters is  $7.03 \times 10^6$  and the GFLOPs is 15.8. It can be seen that after using ShuffleNetV2 and CBAM, the mAP is improved by 1.7% compared to YOLOv5s, the model size is reduced from 14.08MB to 6.69MB, and the GFLOPs are reduced from 15.8 to 5.8; after using the Ghost\_SPPFCSPC\_group structure, the mAP is again improved by 1.5%, with a slight increase in Model Size and GFLOPs; with the use of deep separable convolution, the mAP is 96.9%, Model Size is again reduced to 6.64M, and GFLOPs are 5.5. compared to the original network. mAP is improved by 2.8%, Model Size is reduced by 52.8%, and GFLOPs are reduced by 65.6%. The results show that improvements to YOLOv5 are necessary everywhere.

#### F. Test Results of Defect Detection

Fig. 10 shows the schematic diagram of four kinds of defect detection in aluminum profiles. The non-maximum suppression (NMS) is used in the prediction as the post-processing method. The confidence was set to 0.5 and the IoU was set to 0.6.

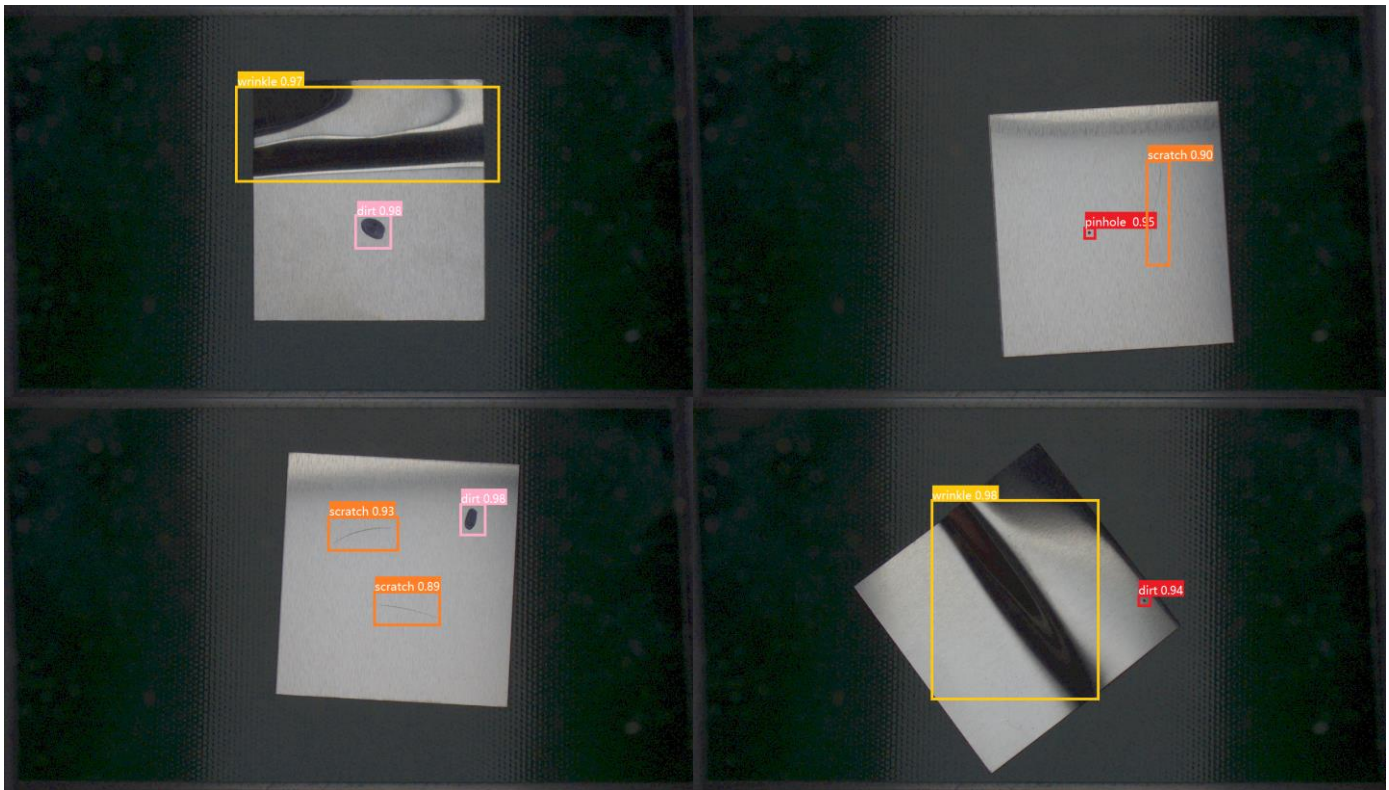


Fig. 10. Partial detection images for aluminum profiles defect detection.

The obtained LAD-YOLO P-R curve for defect detection is shown in Fig. 11, and it can be seen that the AP of pinholes is 98.8%, the AP of dirt is 99.3%, the AP of scratches is 91.1%, and the AP of wrinkles is 98.2%. Except for scratches, all other types of defects have an AP of 98% or more. The poor detection of scratches is due to its high defect precision requirements and susceptibility to environmental influences.

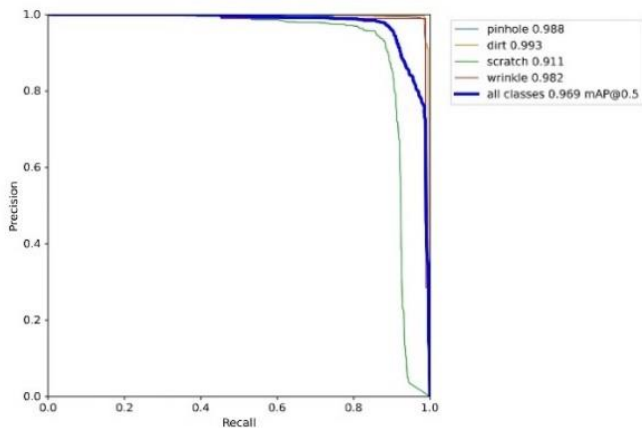


Fig. 11. LAD-YOLO P-R curve.

The accuracy of detecting various defects under different methods is plotted in Fig. 12. It is evident from the figure that LAD-YOLO exhibits improvements in accuracy for different defect types compared to other methods. Specifically, LAD-YOLO shows an increase in accuracy of 4.6% for pinhole

detection, 2.3% for dirt detection, 3.2% for scratch detection, and 0.9% for wrinkle detection when compared to YOLOv5s.

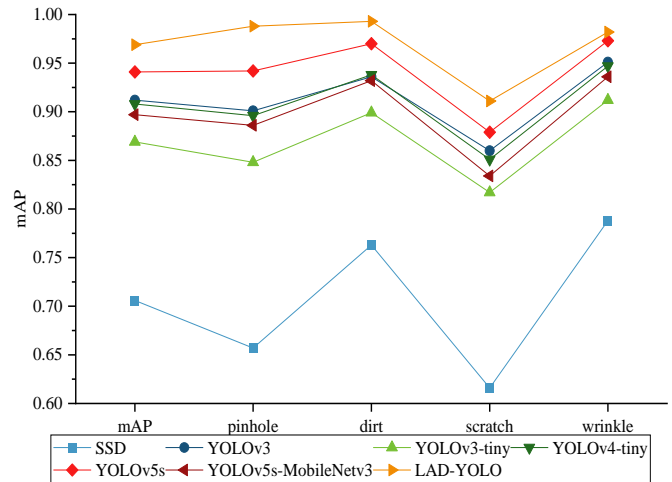


Fig. 12. Accuracy of defects in each category under different methods.

The results demonstrate that LAD-YOLO achieves enhanced accuracy across all defect categories, with particularly notable improvements in the detection of small targets, such as pinholes and scratches.

#### IV. CONCLUSION

We propose a lightweight aluminum profiles surface defect detection network, which involves improvements to the backbone and neck layers of YOLOv5. By designing the Ghost\_SPPFCSPC\_group structure with low floating-point

operation and combining it with the ShuffleNetV2 module and the Convolutional Block Attention Module (CBAM) to construct the backbone network, we reduce the model parameters and computation amount while obtaining richer feature information, thus improving the network's ability to detect defects on small-sized targets. By using depthwise separable convolution to replace the standard convolution in the neck layer, the number and size of model parameters are further reduced to improve the network operation speed. The specific experimental results are as follows:

(1) The Model Size of LAD-YOLO is only 6.64MB, which is 52.84% less compared with YOLOv5s; its GFLOPs are only 5.5, which is 65.19% less compared with YOLOv5s. It shows that LAD-YOLO occupies fewer memory resources, which is more helpful to be applied to platforms with scarce computational resources to achieve low-cost aluminum profiles surface defect detection.

(2) The detection accuracy of LAD-YOLO is much higher than that of current detection methods, including SSD, YOLOv3, YOLOv3-tiny, YOLOv4-tiny, YOLOv5s, and YOLOv5s-MobileNetv3, etc. Compared with YOLOv5s, the mAP of LAD-YOLO is 96.9%, an improvement of 2.8%; compared with YOLOv5s-MobileNetv3, the accuracy is improved by 7.2%. The results indicate that the LAD-YOLO network not only achieves model lightweight but also shows an improvement in accuracy.

In the forthcoming phases, we intend to augment the variety of defects within our dataset, thereby enhancing its diversity. Furthermore, to bolster the model's resilience and versatility in real-world scenarios, we will acquire images portraying authentic situations characterized by uneven lighting, occlusions, and other intricacies. This approach aims to further elevate the model's overall performance.

#### ACKNOWLEDGMENT

This research was funded by the Postgraduate Scientific Research Innovation Project of Changsha University of Science & Technology, Grant Number CXCLY2022141, the Open Research Fund of Hunan Provincial Key Laboratory of Flexible Electronic Materials Genome Engineering, Grant Number 202019, the Open Research Fund of the Hunan Province Higher Education Key Laboratory of Modeling and Monitoring on the Near-Earth Electromagnetic Environments, Grant Number N202107, and the Postgraduate Scientific Research Innovation Project of Hunan Province, Grant Number CX20200896.

#### REFERENCES

- [1] R. Girshick, J. Donahue, T. Darrell, J. Malik; "Rich feature hierarchies for accurate object detection and semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition. 2014.
- [2] R. Girshick, "Fast R-CNN," in Proceedings of the IEEE international conference on computer vision, 2015, pp. 1440-1448.
- [3] 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.
- [4] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in Proceedings of the IEEE international conference on computer vision, 2017, pp. 2961-2969.

- [5] Wei, Liu. "SSD: Single shot multibox detector In Computer vision ECCV2016 14th European conference proceedings Part I (eds Leibe, B., Matas, J., Sebe, N. & Welling, M). 21-37." (2016).
- [6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proceedings of the IEEE Conference on computer vision and pattern recognition, 2016, pp. 779-788.
- [7] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 7263-7271.
- [8] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," arXiv preprint arXiv:1804.02767, 2018.
- [9] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," arXiv preprint arXiv:2004.10934, 2020.
- [10] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "Yolox: Exceeding yolo series in 2021," arXiv preprint arXiv:2107.08430, 2021.
- [11] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 7464-7475.
- [12] Li Q H, Liu D. Aluminum plate surface defects classification based on the BP neural network [J].Applied Mechanics and Materials, 2015.734:543-547.
- [13] Ferguson M K, Ronay A K, Lee Y T, et al. Detection and segmentation of manufacturing defects with convolutional neural networks and transfer learning [J]. Smart and sustainable manufacturing systems. 2018, 2.
- [14] Song L, Lin W, Yang Y, et al. Weak micro-scratch detection based on deep convolutional neural network[J].IEEE Access.2019,7: 27547-27554.
- [15] Duan C, Zhang T. Two-stream convolutional neural network based on gradient image for aluminum profile surface defects classification and recognition[J]. IEEE Access,2020, 8:172152.172165.
- [16] X. Cheng and J. Yu, "RetinaNet With Difference Channel Attention and Adaptively Spatial Feature Fusion for Steel Surface Defect Detection," in IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-11, 2021.
- [17] W. Zeng, Z. You, M. Huang, Z. Kong, Y. Yu and X. Le, "Steel sheet defect detection based on deep learning method", Proc. 10th Int. Conf. Intell. Control Inf. Process. (ICICIP), pp. 152-157, Dec. 2019.
- [18] X. Jin et al., "DM-RIS: Deep multimodel rail inspection system with improved MRF-GMM and CNN", IEEE Trans. Instrum. Meas., vol. 69, no. 4, pp. 1051-1065, Apr. 2020.
- [19] X. Chen and H. Zhang, "Rail Surface Defects Detection Based on Faster R-CNN," 2020 International Conference on Artificial Intelligence and Electromechanical Automation (AIEA), Tianjin, China, 2020, pp. 819-822.
- [20] X. Wei, Z. Yang, Y. Liu, D. Wei, L. Jia, Y. Li. Railway track fastener defect detection based on image processing and deep learning techniques: a comparative study. Eng. Appl. Artif. Intell., 80 (2019), pp. 66-81.
- [21] J. Chen, Z. Liu, H. Wang, A. Núñez, Z. Han Automatic defect detection of fasteners on the catenary support device using deep convolutional neural network. IEEE Trans. Instrum. Meas., 67 (2) (2018), pp. 257-269.
- [22] Zhao H L, Yang Z F, LI J. Detection of metal surface defects based on YOLOV4 algorithm [J]. Journal of Physics: Conference Series, 2021, 1907 (1):12-43.
- [23] Wang, T.; Su, J.; Xu, C.; Zhang, Y. An Intelligent Method for Detecting Surface Defects in Aluminium Profiles Based on the Improved YOLOv5 Algorithm. Electronics 2022, 11, 2304.
- [24] Iandola F N, Han S, Moskewicz M W, et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and < 0.5 MB model size[J]. arXiv preprint arXiv:1602.07360, 2016.
- [25] A. G. Howard et al., "Mobilenets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861, 2017.

- [26] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in Proceedings of the IEEE Conference on computer vision and pattern recognition, 2018, pp. 4510-4520.
- [27] A. Howard, M. Sandler, G. Chu, L. Chen, B. Chen, et al., "Searching for mobilenetv3," in Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 1314-1324.
- [28] X. Zhang, X. Zhou, M. Lin, and J. Sun, "Shufflenet: An extremely efficient convolutional neural network for mobile devices," in Proceedings of the IEEE Conference on computer vision and pattern recognition, 2018, pp. 6848-6856.
- [29] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "Shufflenet v2: Practical guidelines for efficient CNN architecture design," in Proceedings of the European Conference on computer vision (ECCV), 2018, pp. 116-131.
- [30] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "Ghostnet: More features from cheap operations," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 1580-1589.
- [31] Y. Li, Z. Han, H. Xu, L. Liu, X. Li, K. Zhang. Yolov3-lite: a lightweight crack detection network for aircraft structure based on depthwise separable convolutions Appl. Sci.-Basel, 9 (18) (2019)
- [32] D. Xiao, F. Shan, Z. Li, B. T. Le, X. Liu, and X. Li, "A Target Detection Model Based on Improved Tiny-Yolov3 Under the Environment of Mining Truck," in IEEE Access, vol. 7, pp. 123757-123764, 2019.
- [33] H. Zhang, Y. Song, H. Zhong, L. Liu, et al., "MRSDI-CNN: Multi-Model Rail Surface Defect Inspection System Based on Convolutional Neural Networks," in IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 8, pp. 11162-11177, Aug. 2022.
- [34] C. Wang, M. Sun, Y. Cao, K. He, et al. Lightweight Network-Based Surface Defect Detection Method for Steel Plates[J]. Sustainability, 2023, 15(4): 3733.
- [35] Z. Ma, Y. Li, M. Huang, Q. Huang, J. Cheng, et al. Automated real-time detection of surface defects in manufacturing processes of aluminum alloy strip using a lightweight network architecture. J Intell Manuf 34, 2431–2447 (2023).



# Improved YOLO-X Model for Tomato Disease Severity Detection using Field Dataset

Rajasree R, C Beulah Christalin Latha

Department of Digital Sciences, Karunya Institute of Technology and Sciences, Coimbatore, India

**Abstract**—In the past decade, the field of automatic plant disease detection has undergone significant complexity. Advancements in convolutional neural network in deep learning have enabled the rapid and precise detection of ailments, facilitated the development of effective treatments and ultimately led to higher crop yields. One of the most challenging scenarios in plant disease occurs when multiple diseases manifest on a single leaf, exacerbating the difficulty of diagnosis due to overlapping symptoms. This study addresses these challenges by employing an enhanced YOLO-X model for detection tomato leaf diseases. The technique presented here enhances the Spatial Pyramid Pooling layer in order to extract valuable features from training data of various sizes more efficiently. We were able to increase the model's ability to identify a broader spectrum of disease symptoms by concatenating variables from multiple layers and varying sizes. In addition, we incorporate a large number of connections to increase the generalizability of the design. The application of an IoU-based (Intersection over Union) regression loss function increases the convergence of the network and the precision of the detection. For experimentation, we created a customized dataset consisting of 1220 tomato plant leaf images from various farms in Southern part of India, encompassing overlapping diseases and varying degrees of severity. The dataset includes images of healthy leaves as well as different severity levels of tomato leaf curl and tomato leaf mold stress on a single leaf. Our suggested improved SPP-based YOLO-X model beats the original YOLO-X model, according to experimental findings, which show an improvement in test dataset accuracy and a 73.42% mean Average Precision on field-collected dataset.

**Keywords**—Convolutional neural network; deep learning; object classification; plant disease detection; spatial pyramid pooling; YOLOX

## I. INTRODUCTION

In India, tomatoes are a significant economical crop that is produced on 15% of the nation's total cultivated land. A significant portion of the global textile economy is contributed by the nation's tomato production and export, in addition to its local consumption. The crop is afflicted with several diseases during the course of its existence. A leaf might sometimes have many diseases, some of which have similar symptoms. Even an experienced pathologist may make mistakes when evaluating disease severity signs and the presence of numerous stressors. Precision farming practices have undergone a revolution with the development of artificial intelligence and computer vision technology. In plant disease detection systems, a number of machine learning and deep learning models have shown outstanding performance [1] on field-collected or publicly accessible plant disease datasets, some researchers have combined deep learning-based feature extraction and

classification tasks with transfer learning [8]. In order to propose the use of pesticides or other preventative measures and achieve near-ideal performance in recognizing diseases signs automatically, several research have been conducted. Well-known deep learning architectures such as region-based convolutional networks [2], single shot detectors [3], and region proposal networks [4] have been employed in the area of plant leaf disease detection, with major alterations happening during the preceding few years [37, 38, 41]. Almost all previous studies either used the well-known PlantVillage public dataset [3] or their own datasets collected in the field [5], [6]. But only a small number of studies have looked at the stages of disease growth and the chance that many living and nonliving things can attack a plant leaf at the same time. In these situations, it is hard for both human and automatic monitoring systems to figure out the type of infection and the exact area of sickness signs.

In this study, we describe a YOLO-X-s based detecting system that uses a modified Spatial Pyramid Block to combine fine spatial data with local features to find sickness phases and split diseases with symptoms that overlap. We made the spatial pyramid pooling block better by putting together feature maps at low-level scales. This helped us solve the problem more accurately. The original size feature vector was added to improve the quality of the features. The recognition performance got even better when the Alpha IoU regression loss function was used [36].

### A. Contributions to the Research

- A better YOLO-Xs model with a modified Pyramid pooling module (SPP) layer is given so that many diseases on a single leaf plant can be found. It collects location information at local, multi-scale levels to get the information it needs more quickly.
- To improve generalization and convergence, we used Alpha IoU (Intersection over Union) loss as the bounding box regression for multiple disease localization when multiple diseases showed up on the same plant leaf.
- With the help of enhancement, a group of unique shots from a tomato field are shown. The photos show how diseases spread and how many different diseases can be found on a single leaf.

The paper has been structured in the following manner: The Section II provides a summary of the existing literature. The Section III describes the proposed methodology. The Section IV presents research outcomes, comparisons to existing

methods and outcomes, and discussions of future research opportunities and limitations of the study. Section V provides a conclusion for this proposed research.

## II. LITERATURE REVIEW

Plant leaf diseases may be identified using computer vision-based methods for (1) detection and (2) identification. Both of the methodologies used in the area are prevalent in research literature released in the last ten years.

Highlights of various modern and cutting-edge techniques for identifying and detecting plant leaf diseases literatures are included in Table I. In this part, these methods will be thoroughly reviewed in relation to: (1) the kind of application targeted, (2) the methodology employed, (3) the contribution. The usage of deep learning in this field of study has increased during the previous several years. Transfer learning and data enrichment have made it easier to use deep learning models on a variety of devices, such as central processing units (CPUs) and graphics processing units (GPUs). In the study done by a research [10], the MobileNet v2 model was trained by adding more color space data in a few different ways. Transfer learning and data enrichment have made it easier to run deep learning models on CPUs, GPUs, and other types of computer systems. This is because these two methods have been put together. In the study done by a researcher [10], color space data addition methods were used to train the MobileNet v2 model. To compare the effectiveness of the classification, the scientists trained the model using images of cassava leaf disease of various quality levels. According to a study, low-quality images cause the classification accuracy to decrease. The diseased region has also been identified using high-quality images and a color difference, according to authors [11]. Using advanced machine learning classifiers, such as the bagging tree ensemble, it is now possible to identify sick regions based on color and textual information with an overall accuracy of 99%. Plant disease monitoring systems that use computer vision are meant to automatically find and identify the part of a plant that is sick. Because of this, these systems use customizable deep-learning meta-architectures that have been used in the related study.

The earliest deployed deep learning algorithms were region-based convolutional neural networks (RCNNs), and their purpose was to recognize objects in general [3]. In line with Fast Convolutional Neural Network [12], Faster Region based CNNs, and R-FCNNs [3]. Segmentation and noise reduction operations were carried out using the OTSU algorithm and multilayer median filters, respectively. By using a two-stage detector, the technique achieves an inference time of 0.52 s. A research study [14] employed a comparable two-stage detector, Mask RCNN, to identify sick regions after contrast stretching. CNN performs feature extraction of improved areas, which was afterwards categorized. After applying entropy to choose the best features, accuracy was improved. Another important work [15] identified various rice diseases by using a quicker R-CNN with a reinforced backbone to analyze the still pictures of the rice. In terms of recall and accuracy, the upgraded two-stage model performs better than earlier models such as YOLO-V3, while having longer detection times. The model presented in [16] significantly

improved mAP by using a superior anchor box method that was based on a more efficient RCNN model for weed detection.

With single shot detection (SSD), need to make a region proposal network in order to get the best total speed and a faster inference time. But you can get both without making a region proposal network. They use certain boxes to figure out how likely it is that a certain item will be in a picture. This improved model, which used the Inception module and Rainbow union, was used to get information about features and make it easier to find ill spots on apples [17]. Both the VGG and the origin module were added to the model so that it could diagnose diseases with a mAP of 78.8%. YOLO (You Only Look Once) [18] models, on the other hand, are single-stage detectors that can find and label objects in a picture with just one forward spread. The name of these models comes from the saying "you only look once." It initially divides a picture into a grid to begin the detection process, and for each bounding box, it then forecasts the likelihood of an item and its class. In order to identify tomato diseases in difficult background settings, Wang et al. [19] utilized a similar YOLO architecture and added the DenseNet block for feature extraction. Additional information is provided in Table I of the relevant work comparison. The YOLO-V5 model was effective in identifying bacterial leaf spots. Additionally, the outcomes were contrasted with those of YOLO-V3, YOLO-V4, Single Shot Detector algorithm, and other two-stage detectors techniques [20].

A new upgrade to the YOLO series, called YOLO-X [26], has made a considerable improvement in the field of object identification. To improve feature extraction, YOLO-X makes use of the YOLO-V3 with a Darknet architecture added with SPP Layer baseline structure. To diagnose kiwi leaf diseases, DeepLabV3 and UNet are used in conjunction with YOLO-X to separate the leaves from the complicated backdrop. The mix of Cross Stage Partial Network and sigmoid activation function in the approach of finding colon cancer has led to a better version of the YOLO v3 algorithm.

The model is strengthened for real-time polyp identification by using the CIU loss function [28]. The YOLO-V5 model was used in a research work [22] to accurately detect plant disease. Using the SE module and Involution Bottleneck, the accuracy and number of parameters were improved. The researchers must overcome difficulties with disease development and detecting many lesion areas in a single frame. Multiple diseases may be difficult to identify, both manually and when using artificial image processing algorithms [29], since their locations and symptoms tend to overlap. The EfficientNet model, which has a classification accuracy of 96%, was successfully used to address the plant disease detection problem on a single cucumber leaf [25] and other notable research [42] explains the hyper-tuned efficientdet architecture for object detection. The authors utilized a Ranger optimizer to find symptoms that seemed to be related. Although SSD algorithm with respectable object identification and detection performance have been employed in the methodologies from the literature (included in Table I).

TABLE I. SELECTED STUDIES FROM THE LITERATURE ON PLANT DISEASE DETECTION

Reference	Application Targeted	Methodology Employed	Contribution
[7]	Automatic detection of plant diseases	Classic machine learning	Utilization of various classic machine learning approaches
[13]	Rice disease identification	Faster region-based model + K-means clustering	Identification of rice diseases using a two-stage detector with clustering
[17]	Identification of apple sick patches	YOLO-V5 model with Inception module and Rainbow concatenation	Enhanced feature extraction and improved identification of apple sick patches
[19]	Tomato disease identification	YOLO architecture with DenseNet block	Utilization of YOLO architecture for identifying tomato diseases
[22]	Plant disease detection	YOLO-V5 model with SE module and Involution Bottleneck	Accurate detection of plant diseases with improved accuracy and parameters
[23]	Segmentation of sick lesions on paddy leaves	CNN and YOLO	Segmentation of sick lesions with refined model parameters
[24]	Classification of severity of strawberry leaf disease	Faster RCNN and Siamese network	Identification of leaf position and estimation of severity
[42]	Object detection	Efficientdet	Identification of hyper-tuned efficientdet architecture for object detection

The majority of these investigations have focused on classifying lesions or locating locations. The difficulties associated with accurately distinguishing distinct phases of disease severity and a multitude of stresses on a single leaf are presently the ones that have received the least amount of attention. In order to identify overlapping and overlapping plant leaf diseases, we suggest in this study an updated YOLO-X model tailored specifically for the application.

### III. METHODOLOGY

The suggested technique is fully explained in this part and will be utilized to handle the issues of (1) leaf disease progression symptoms and (2) the presence of several diseases on a single tomato leaf. The supplied data is initially pre-processed to remove extraneous background information and guarantee class balance. The suggested deep learning model's specifics are then discussed, with a focus on the two crucial instances indicated above.

#### A. Customized Dataset Creation

Tomatoes are grown in different parts of Southern India from March to May. For the purpose of this study, images were taken from various tomato fields in Tamil Nadu and Kerala. No extra pesticides or fertilizers were used to maintain the conditions that allowed the infections to flourish. The pictures were taken at various angles using a smart phone (Apple iPhone X) that was held 15 to 25 cm away from the tomato

leaf. Original image size was 1125 x 2436 pixels with 19.5:9 ratio. Images were taken in the morning, with varied lighting conditions. The smart phone's focus was changed to portrait mode and zoom mode to capture one leaf. Seven different categories of tomato curl severity and disease coexistence images were created from the data.

Fig. 1 displays a few examples of images extracted from the dataset which was collected from the fields. In India, the ailment known as "tomato curl virus" is widespread. Stage 1 of tomato leaf curl's early signs is described as leaf chlorosis. Within two to three weeks, the symptoms become worse as the leaf's veins start to thicken and darken. From the underside of the leaf, the thickening of the veins is plainly visible. The second stage of the tomato leaf disease is distinguished by the subsequent curling of the leaf margins. Three weeks later, the leaves begin to develop dark black specks that prevent the virus from spreading. The pathologists determine that it is tomato leaf mold brought on by the aphid infestation. The curl virus and leaf mold infect almost all of the nearby plants.



Fig. 1. Sample dataset images collected from the tomato field.

#### B. Pre-processing the dataset images

Since leaf photographs were taken in actual field settings, they included background details like dirt and tree branches, among other things. The work of the suggested detection system is anticipated to be made simpler by extracting just the leaf region containing one or more diseases and deleting the background noises [30]. There are many methods in place to eliminate extraneous background details from real-world situations [9]. Grab Cut technique [31] is a quick and effective machine learning-based approach that can eliminate unwanted background data with little human adjustments. On the basis of graph cuts, the backdrop is removed. According to a user-provided window, anything beyond the window is taken into account to be the clear backdrop, however within the window, both the foreground and the background may exist.

The procedure is repeated until convergence in order to fine-tune the background removal job. On the hardware utilized for testing, background removal was reported to take an average of 4.12 seconds per picture. The algorithm's ability

to eliminate background data outside of the infected leaf was determined to be effective. After the backgrounds have been removed, the images are annotated before being subjected to image augmentation methods including flipping, rotating, and brightness boosting.

### C. Annotation and Labelling

The Roboflow annotation tool was used to manually design the ground truth boxes and labeling the healthy and diseased tomato leaf images. According to Fig. 2 it shows the experiments done for the annotation and labeling, sample taken from the roboflow tool. Each box's enclosing box coordinates, height, breadth, and class name are listed in Fig. 3. When an image is included into the model for learning, testing, or assessment, an accompanying XML file is included [17].

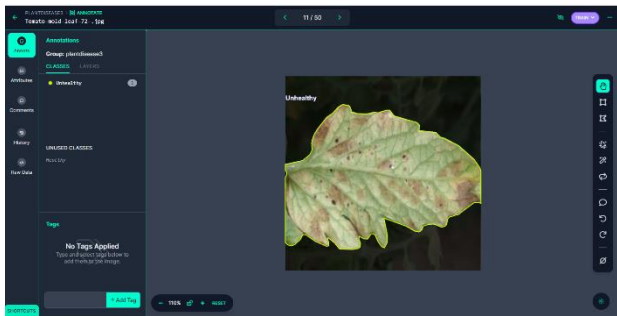


Fig. 2. Tomato leaf image annotation with bounding box created using roboflow tool.

```
{  
  "boxes": [  
    {  
      "label": "sooty",  
      "x": 272,  
      "y": 164,  
      "width": 93,  
      "height": 62  
    },  
    {  
      "label": "curl_stage2",  
      "x": 130.5,  
      "y": 184.5,  
      "width": 57,  
      "height": 63.5  
    },  
    {  
      "label": "sooty",  
      "x": 214,  
      "y": 284.5,  
      "width": 71,  
      "height": 75  
    }  
  ]  
}
```

Fig. 3. Annotation details of the xml file taken from the roboflow tool.

Images captured at different angles and sizes are included in the dataset used for the tests. In order to maintain the bounding boxes on the item of interest regardless of any augmentation step applied to it subsequently, we executed an auto-orient procedure. As often requested by YOLO, the photographs were also scaled using a normalization set at a resolution of 416X416. When working with an unbalanced dataset, we prefer to execute the augmentation techniques to minimize over-fitting. Proposed study used augmentation

techniques includes flip horizontal, rotation, and brightness (25%). The addition of these augmentations will increase the dataset's size as well as the variety of the photographs taken in various lighting situations. The dataset is augmented at random, increasing the size to 1, 112. Roboflow data services were used for each of these pre-processing phases.

### D. Hardware and Software Configuration

Every experiment was run on Windows 7 system using an Intel i3 processor. The necessary repositories and libraries were installed before configuring the experimental environment. Pre-trained weights for the YOLO-X model were downloaded when the dataset was uploaded to the drive. Experiment was performed using google collab notebook environment. With the help of the hyper-parameters using 100 epochs, our suggested model was trained. The best weights produced after training were kept and then utilized to assess the effectiveness of our suggested model on test images with a batch size of 32. In order to conduct the experiment, 640 × 640 resolution images from the dataset were used. The dataset has been split automatically with the code on the collab on the basis of 80% of the images were utilized for training and testing, while the other 20% were used for validation process.

### E. Enhanced YOLO-X Proposed Model for Tomato Leaf Disease Detection

YOLO- X deep learning model is a single stage object detection technique that differs from YOLO-v3. It derives from DarkNet53 architecture. Beginning with a 1 X 1 convolutional layer, the feature channel for each level of FPN features is reduced to 256. Then, for the classification and regression tasks, respectively, we add two parallel branches with two 3X3 convolution layers. The anchor-free detector YOLO-X [26] has shown exceptional speed and accuracy performance. To improve convergence while the system was being trained, the head of YOLO-X was cut off from the original detector [32]. Because to the implementation's anchor-free design, the overall number of trainable parameters has been significantly decreased.

SimOTA [26], which reduces the amount of time required for training and aids in the solution of the Optimal Transport (OT) issue [21], is also used to enhance the label assignment approach.

For the application that is covered in this article, we build a more sophisticated YOLO-X model. The CSP darknet serves as the foundation for feature extraction in this model. In order to accomplish better feature fusion throughout the classification and regression tasks, the feature layers are first up-sampled for feature fusion and then down-sampled for classification [27]. One of the crucial elements that must exist for good feature extraction is the focus module. The input photographs are split into four pieces, then concatenated in order to preserve information about the features of the objects. This makes it possible to observe the characteristics more clearly. The Bottleneck CSP layer, which comes after the top layer, is where the deep features are recovered with more accuracy. Convolution, related batch normalization, and activation processes are performed on the feature maps. To learn about overlapping and mild symptoms, advanced data augmentation methods like mosaic and mix-up are utilized

during training [33]. The non-maximal suppression (NMS) strategy prevents the chance of multiple detections happening at the same time.

A definition for the modeling error that arises between the anticipated class and the ground truth is Binary Cross Entropy (BCE) loss [34] with logits. A sigmoid activation function is utilized to eliminate all accurate predictions [26]. The bounding box's coordinates (x, y, w, and h) are predicted in the regression branch's output.

In order to forecast bounding box outputs, YOLO-X employs the IoU metric and compares its predictions to the actual data.

#### F. YOLO X Model Working Environment with Bounding Box

The localization of objects and their categorization are two processes that are critically important for applications based on computer vision. How precisely a machine learning model can pinpoint an object's placement inside a scene or image is determined by the loss function [35]. This is why conventional single-stage and two-stage detectors are developed using the bounding box regression approach.

Vanishing gradients provide a challenge because they cause IoU losses, which prohibit the model becoming convergent. The basis of the problem is that the predicted boxes do not precisely overlap the ground truth boxes, leading to inaccurate findings. To increase the precision with which the objects were tracked, it was chosen to give a number of enhanced bounding box losses depending on a number of parameters. The Generalized Intersection over Union, the Distance Intersection over Union, the Complete Intersection over Union, and the Efficient Intersection over Union (GIoU, DIoU, CIoU, EIoU) are acronyms for these intersections. The amount of overlap between the target and the anchor boxes is represented by these loss functions via the use of several metrics.

### IV. RESULTS

Training effectiveness was assessed based on increased convergence speed and detection performance of overlapping illness symptoms and severity classes using datasets collected from the field. For detection models, the Mean Average Accuracy (mAP) statistic is often used. This statistic indicates the accuracy for all classes when the Intersection over Union (IoU) criteria is set to 50%. The IoU threshold was held constant at 0.5 while calculating the mAP score. Similar levels of training accuracy were originally shown by the default YOLO-X model, but it was unable to converge in the last 30 epochs of training. As illustrated in Fig. 4, this resulted in the default YOLO-X model's accuracy being lower (69.90%) than the suggested enhanced YOLO-X model's accuracy (73.42%).

A SE block was added to the SPP module of the standard YOLO-X in an effort to boost detection performance. [39] To achieve this, many tests were conducted. The YOLO-X-SE model's training performance was plainly overfit in the last 20 training epochs, which decreased inference speed. Several unique classes were misclassified, despite the fact that our model's inference time is a little faster than the default YOLO-X. The model's validation and test results were assessed over a range of IOU losses in order to attain the best degree of

convergence and mAP performance. The model performed noticeably better in terms of the results of the enhanced SPP block compared to vanilla-IoU and other regression techniques. For the localisation of overlapping medical symptoms, this was crucial.

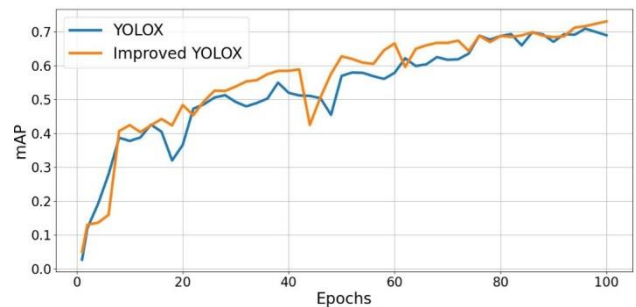


Fig. 4. YOLOX vs improved YOLOX model performance (mAP).

The best weights are chosen when the training process is finished, and the model's performance on the test dataset is assessed using these weights. The regression will come to an end once the item included by the bounding box has been located. The BCE loss may be used to assess a bounding box's capacity to hold an item. The disease's ability to exist inside the anticipated bounding box is assessed using the confidence score. We conducted a number of tests to determine the ideal threshold value in order to increase the detection mAP score. This figure shouldn't be too high or low in order to avoid false positives and genuine predictions, respectively. The degree of confidence will stay at 0.25 after an analysis of the test data. Each projected bounding box in the test photos has a corresponding confidence level, which may be used to represent the detection performance of the test dataset as determined by Improved YOLO-X. The result offers proof that strengthens confidence in the localization and classification procedures. Fig. 5 represent the loss graph of YOLOX and improved YOLOX model in each epoch.

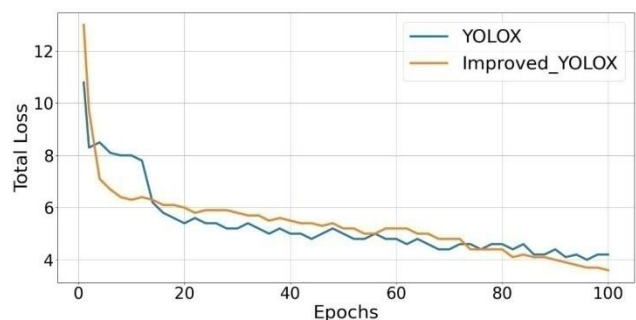


Fig. 5. Graph depicts the YOLOX vs improved YOLOX loss vs epochs.

The YOLO-X model uses the same basic idea as its predecessor but with an SPP anchor-free system. The decoupled head lowers convergence, as shown by comparing its performance to that of the YOLO-V4 [40] and YOLO-V5 [20] models, both of which were trained and validated using our Tomato severity dataset. The findings are shown in Table II depending on the image size and processing inference time needed for inference. The accuracy of detection has reduced even though YOLO-V4 and YOLO-V5 have inference rates that are noticeably quicker than our model. Both the default

YOLO-X model and our Improved YOLO-X model demonstrate higher convergence performance compared to other models trained on our dataset because the anchor-free method decreases the possibility of complexity and obstacles emerging during training.

TABLE II. RESULT OF MODEL COMPARISON BASED ON TIME ATTRIBUTE

Experimented model	Input Size	Inference Time
YOLO-v4	416X416	20.81
YOLO-v5	416X416	16.01
YOLO-X	640X640	32.08
YOLO-X-SE	640X640	56.42
Enhanced YOLO-X	640X640	32.03

To evaluate how well our suggested YOLO-X model performs in classifying overlapping symptoms and severity phases. Because it was incorrectly categorized with the healthy and leaf curl disease stage-2 class, curl stage-1 was recognized with substantially lower average precision. The overlapping signs of leaf mold and curl stage-2 were also more accurately picked up by our upgraded YOLO-X model. As the class does not visually resemble any other classes and there are not many problematic images in our test dataset, the models successfully identify leaf stress.

#### A. Comparison with Other Existing Models

Here, we contrast the effectiveness of our upgraded YOLO-X model with that of current state-of-the-art models. The performance of our proposed model is contrasted with that of cutting-edge models in Tableau. We trained the YOLO-V4, YOLO-V5, YOLO-V7, Efficientdet, and YOLO-X models on our tomato severity dataset for 100 iterations using the default code settings. The mAP scores are noticeably lower than those anticipated by the YOLO-X model, as seen in Table III. We created the YOLO-X- lite model, a YOLO-X version that works well for edge computing. An SPP block that has been tailored for embedded device operation is included in the model. However, it did not provide the expected results for a specific dataset. The advantages of applying Spatial Pyramid Pooling findings into mAP analysis are shown in Table IV. It indicates that when SPP with 3, 5, 7, and 9 connections is taken into account, our optimized YOLO-X model performs better.

TABLE III. RESULT ANALYSIS OF THE PROPOSED MODEL IN COMPARISON WITH OTHER STATE OF HEART ALGORITHMS

Experimented model	mAP in Healthy	mAP in Leaf Curl Disease - Stage 1	mAP Leaf Curl Disease - Stage 2	mAP in leaf mold
YOLO-v4	27.87%	56.61%	40.12%	54.01%
YOLO-v5	65.91%	29.21%	47.72%	53%
YOLO-X	64.21%	49.58%	61.23%	62.31%
YOLO-X-SE	56.01%	63%	64.02%	59.82%
Enhanced YOLO-X	62.37%	62.32%	65.72%	75.02%

TABLE IV. AN ANALYSIS OF EVALUATION METRICS USING MEAN AVERAGE PRECISION WITH SPATIAL PYRAMID POOLING

Spatial Pyramid Pooling (5,9,14)	Spatial Pyramid Pooling (3,5,7,9)	Skip Connections	mAP
Yes	No	No	69.90 %
No	Yes	No	67.42%
No	Yes	Yes	71.31%
No	Yes	Yes	73.42%

#### B. Future Work and Research Opportunities

Multiple stresses on the host and multiple disease stages on the leaf are typical outdoor conditions. For these two scenarios, this research suggests a deep learning-based solution that has been put to the test on a large dataset. The scope of the proposed study is restricted to assessing the severity of diseases only on the leaves of tomato plants. Future study might potentially prioritize the study of disease severity detection in various parts of the plant. In order for our model to be applicable in real-world scenarios, its reliability is of utmost importance. To achieve this, the dataset will be continuously enhanced by incorporating new photographs depicting various ailments and harvests.

- It is strongly advised to subject the recommended model to additional training and testing, specifically using individual leaves placed against a clean background.
- To enhance its practicality and efficiency, we intend to allocate a greater financial investment towards obtaining more field samples.
- Furthermore, to expedite the training and testing processes on state-of-the-art technology, we will prioritize designing the model to be as lightweight as possible.

#### V. CONCLUSION

The suggested study offers a framework for classifying symptoms of a particular disease on tomato plants according to increasing severity. The detection of many illnesses that are present on a single leaf may also be done using this approach. We have suggested a YOLO-X-based model with an enhanced Spatial Pyramid Pooling block to achieve this. Various pooling rates were used to aggregate multi-scale characteristics. Remaining links were added to better preserve spatial information. With the use of this model, we were able to identify diseases symptoms that were similar and overlapped more accurately. The suggested model outperformed the default YOLO-X by 3.27 percent according to experimental findings, achieving mAP scores of 73.42% and 72.31% using training dataset and testing dataset respectively. Additionally, Curl stage-2 and Leaf mold yielded the greatest results, with average precisions of 65.76% and 74.02%, respectively, for overlapping and co-existing classes.

REFERENCES

- [1] J. G. A. Barbedo, "Factors influencing the use of deep learning for plant disease recognition," *Biosyst. Eng.*, vol. 172, pp. 84–91, Aug. 2018.
- [2] Z. Lin, S. Mu, F. Huang, K. A. Mateen, M. Wang, W. Gao, and J. Jia, "A unified matrix-based convolutional neural network for fine-grained image classification of wheat leaf diseases," *IEEE Access*, vol. 7, pp. 11570–11590, 2019.
- [3] M. H. Saleem, S. Khanchi, J. Potgieter, and K. M. Arif, "Image-based plant disease identification by deep learning meta-architectures," *Plants*, vol. 9, no. 11, p. 1451, Oct. 2020.
- [4] R. Wang, L. Jiao, C. Xie, P. Chen, J. Du, and R. Li, "S-RPN: Sampling-balanced region proposal network for small crop pest detection," *Comput. Electron. Agricult.*, vol. 187, Aug. 2021, Art. no. 106290.
- [5] D. Jiang, G. Li, C. Tan, L. Huang, Y. Sun, and J. Kong, "Semantic segmentation for multiscale target based on object recognition using the improved Faster-RCNN model," *Future Gener. Comput. Syst.*, vol. 123, pp. 94–104, Oct. 2021.
- [6] A. Fuentes, S. Yoon, S. C. Kim, and D. S. Park, "A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition," *Sensors*, vol. 17, no. 9, p. 2022, 2017.
- [7] N. Kundu, G. Rani, V. S. Dhaka, K. Gupta, S. C. Nayak, S. Verma, M. F. Ijaz, and M. Woźniak, "IoT and interpretable machine learning based framework for disease prediction in pearl millet," *Sensors*, vol. 21, no. 16, p. 5386, Aug. 2021.
- [8] Rajasree, R., Latha, C.B.C., Paul, S. (2022). Application of Transfer Learning with a Fine-tuned ResNet-152 for Evaluation of Disease Severity in Tomato Plants. In: Shakya, S., Ntalianis, K., Kamel, K.A. (eds) Mobile Computing and Sustainable Informatics. Lecture Notes on Data Engineering and Communications Technologies, vol. 126. Springer, Singapore. [https://doi.org/10.1007/978-981-19-2069-1\\_48](https://doi.org/10.1007/978-981-19-2069-1_48).
- [9] B. M. Patil and V. Burkpalli, "Segmentation of tomato leaf images using a modified chan vese method," *Multimedia Tools Appl.*, vol. 81, no. 11, pp. 15419–15437, May 2022.
- [10] O. O. Abayomi-Alli, R. Damaševičius, S. Misra, and R. Maskeliūnas, "Cassava disease recognition from low-quality images using enhanced data augmentation model and deep learning," *Exp. Syst.*, vol. 38, no. 7, 2021, Art. no. e12746.
- [11] A. Almadhor, H. T. Rauf, M. I. U. Lali, R. Damaševičius, B. Alouffi, and A. Alharbi, "AI-driven framework for recognition of guava plant diseases through machine learning from DSLR camera sensor based high resolution imagery," *Sensors*, vol. 21, no. 11, p. 3830, 2021.
- [12] A. Pramanik, S. K. Pal, J. Maiti, and P. Mitra, "Granulated RCNN and multi-class deep SORT for multi-object detection and tracking," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 6, no. 1, pp. 171–181, Feb. 2022.
- [13] G. Zhou, W. Zhang, A. Chen, M. He, and X. Ma, "Rapid detection of Rice disease based on FCM-KM and faster R-CNN fusion," *IEEE Access*, vol. 7, pp. 143190–143206, 2019.
- [14] Z. U. Rehman, M. A. Khan, F. Ahmed, R. Damaševičius, S. R. Naqvi, W. Nisar, and K. Javed, "Recognizing apple leaf diseases using a novel parallel real-time processing framework based on mask RCNN and transfer learning: An application for smart agriculture," *IET Image Processing*, vol. 15, no. 10, pp. 2157–2168, 2021.
- [15] D. Li, R. Wang, C. Xie, L. Liu, J. Zhang, R. Li, F. Wang, M. Zhou, and W. Liu, "A recognition method for Rice plant diseases and pests video detection based on deep convolutional neural network," *Sensors*, vol. 20, no. 3, p. 578, Jan. 2020.
- [16] M. H. Saleem, J. Potgieter, and K. M. Arif, "Weed detection by faster RCNN model: An enhanced anchor box approach," *Agronomy*, vol. 12, no. 7, p. 1580, Jun. 2022.
- [17] P. Jiang, Y. Chen, B. Liu, D. He, and C. Liang, "Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks," *IEEE Access*, vol. 7, pp. 59069–59080, 2019.
- [18] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.
- [19] X. Wang and J. Liu, "Tomato anomalies detection in greenhouse scenarios based on YOLO-dense," *Frontiers Plant Sci.*, vol. 12, p. 533, Apr. 2021.
- [20] M. P. Mathew and T. Y. Mahesh, "Leaf-based disease detection in bell pepper plant using YOLO v5," *Signal, Image Video Process.*, vol. 16, no. 3, pp. 841–847, Apr. 2022.
- [21] J. Yao, Y. Wang, Y. Xiang, J. Yang, Y. Zhu, X. Li, S. Li, J. Zhang, and G. Gong, "Two-stage detection algorithm for kiwifruit leaf diseases based on deep learning," *Plants*, vol. 11, no. 6, p. 768, Mar. 2022.
- [22] Z. Chen, R. Wu, Y. Lin, C. Li, S. Chen, Z. Yuan, S. Chen, and X. Zou, "Plant disease recognition model based on improved YOLO-V5," *Agronomy*, vol. 12, no. 2, p. 365, Jan. 2022.
- [23] G. Ganesan and J. Chinnappan, "Hybridization of ResNet with YOLO classifier for automated paddy leaf disease recognition: An optimized model," *J. Field Robot.*, vol. 39, no. 7, pp. 1085–1109, Oct. 2022.
- [24] J. Pan, L. Xia, Q. Wu, Y. Guo, Y. Chen, and X. Tian, "Automatic strawberry leaf scorch severity estimation via Faster R-CNN and few-shot learning," *Ecol. Informat.*, vol. 70, Sep. 2022, Art. no. 101706.
- [25] P. Zhang, L. Yang, and D. Li, "EfficientNet-B4-Ranger: A novel method for greenhouse cucumber disease recognition under natural complex environment," *Comput. Electron. Agricult.*, vol. 176, Sep. 2020, Art. no. 105652.
- [26] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLO-X: Exceeding YOLO series in 2021," 2021, *arXiv:2107.08430*.
- [27] I. Pacal, A. Karaman, D. Karaboga, B. Akay, A. Basturk, U. Nalbantoglu, and S. Coskun, "An efficient real-time colonic polyp detection with YOLO algorithms trained by using negative samples and large datasets," *Comput. Biol. Med.*, vol. 141, Feb. 2022, Art. no. 105031.
- [28] H. Zhai, J. Cheng, and M. Wang, "Rethink the IoU-based loss functions for bounding box regression," in *Proc. IEEE 9th Joint Int. Inf. Technol. Artif. Intell. Conf. (ITAIC)*, Dec. 2020, pp. 1522–1528.
- [29] S. K. Noon, M. Amjad, M. A. Qureshi, and A. Mannan, "Use of deep learning techniques for identification of plant leaf stresses: A review," *Sustain. Comput., Informat. Syst.*, vol. 28, Dec. 2020, Art. no. 100443.
- [30] R. R. C. B. C. Latha, S. Paul, A. M and A. N, "An optimized Faster R-CNN model for Cassava Brown Streak Disease Classification," 2023 3rd International Conference on Advances in Computing, Communication, Embedded and Secure Systems (ACCESS), Kalady, Ernakulam, India, 2023, pp. 94–100, doi: 10.1109/ACCESS57397.2023.10200536.
- [31] S. Sun, M. Jiang, D. He, Y. Long, and H. Song, "Recognition of green apples in an orchard environment by combining the GrabCut model and Ncut algorithm," *Biosyst. Eng.*, vol. 187, pp. 201–213, Nov. 2019.
- [32] Y. Li, Z. Guo, F. Shuang, M. Zhang, and X. Li, "Key technologies of machine vision for weeding robots: A review and benchmark," *Comput. Electron. Agricult.*, vol. 196, May 2022, Art. no. 106880.
- [33] G. Wang, H. Zheng, and X. Zhang, "A robust checkerboard corner detection method for camera calibration based on improved YOLO-X," *Frontiers Phys.*, vol. 9, p. 828, Feb. 2022.
- [34] U. Ruby and V. Yendapalli, "Binary cross entropy with deep learning technique for image classification," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 4, pp. 5393–5397, Aug. 2020.
- [35] Z. Gevorgyan, "SLoU loss: More powerful learning for bounding box regression," 2022, *arXiv:2205.12740*.
- [36] J. He, S. Erfani, X. Ma, J. Bailey, Y. Chi, and X.-S. Hua, " $\alpha$ -IoU: A family of power intersection over union losses for bounding box regression," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 20230–20242.
- [37] C.-L. Wang, M.-W. Li, Y.-K. Chan, S.-S. Yu, J. H. Ou, C.-Y. Chen, M.-H. Lee, and C.-H. Lin, "Multi-scale features fusion convolutional neural networks for Rice leaf disease identification," *J. Imag. Sci. Technol.*, vol. 66, no. 5, pp. 1–12, 2022.
- [38] T. Wu, Q. Huang, Z. Liu, Y. Wang, and D. Lin, "Distribution-balanced loss for multi-label classification in long-tailed datasets," in *Proc. Eur. Conf. Comput. Vis. Cham, Switzerland: Springer, 2020*, pp. 162–178.
- [39] G. Li, X. Huang, J. Ai, Z. Yi, and W. Xie, "Lemon-YOLO: An efficient object detection method for lemons in the natural environment," *IET Image Process.*, vol. 15, no. 9, pp. 1998–2009, Jul. 2021.

- [40] A. Bochkovskiy, C.-Y. Wang, and H.-Y. Mark Liao, "YOLO-V4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.
- [41] C.-Y. Wang, A. Bochkovskiy, and H.-Y. Mark Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," 2022, *arXiv:2207.02696*.
- [42] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 10781–10790.



# He and She in Video Games: Impact of Gender on Video Game Participation and Perspectives

Deena Alghamdi

Department of Computer Science and Information Systems, Umm Al-Qura University, Makkah, Kingdom of Saudi Arabia

**Abstract**—Playing video games is now considered one of the day-to-day activities of many adolescents and young people. This research studies the gender impact on video game participation and perspectives among college students in the Kingdom of Saudi Arabia (KSA). The data were collected by first conducting discussions involving four focus groups with a total of 26 participants to explore the topic. An online questionnaire was then distributed, and a total of 2,756 responses were received. The analysis of the data shows a clear impact of gender on the playing practices adopted, perceptions towards the pros and cons of video games, and the most used consoles and popular games. However, the practices and perspectives of male and female players did not differ regarding bullying in video games. The findings of this study can advance the understanding of this subject, and game developers who are targeting the KSA game market can use the results as the basis for developing games that are more suitable for the players in that country.

**Keywords**—College students; gender differences; KSA; video games

## I. INTRODUCTION

The Kingdom of Saudi Arabia (KSA) is the nineteenth biggest gaming market in the world and is currently experiencing an enormous (41.1%) year-on-year growth in this sector [1], with 21.1 million gamers in KSA in 2020 [2]. Several studies have discussed the possible impact of violent video games and potential gender-related effects [3] [4] [5]. While [6] argued for the positive effects of video games on well-being greatly depend on the presence of moderation; the aspects involved, such as social aspects, violence, or physical activity; and the motivations behind playing the game. In addition, Comeran-Chueca et al. [7] emphasized on this positive effect when they used active video games (AVG) as an effective strategy to increase energy expenditure in children and adolescents with overweight and obesity when they found that energy expenditure with AVG combined with multi-component exercise was 5.68 kcal/min in boys and 4.66 kcal/min in girls with overweight and obesity.

In the present study, the author explored gender differences in video game-playing among college students in the KSA, players' perspectives of the male and female characters' roles in the video games that they played, and the potential interference of gaming with other aspects of college students' lives. This study aimed to answer the following research questions: (1) What are the main areas that identify the practices and perceptions of players in the KSA? (2) Are there differences between male and female players' perceptions towards video games in the KSA? (3) Does gaming interfere with academic preparation and interpersonal relationships?

In this paper, we illustrate related previous research, describe and discuss our methodological approach, outline the findings about the participants' participation and perspectives regarding video games, and finally discuss the implications of these practices in the KSA.

## II. LITERATURE REVIEW

Previous studies explored and discussed the impact of gender on different aspects of video games such as playing duration, the content of video games, and how this is affecting players' personal life. One study [8] reported that more boys than girls played video games at least once based on two 24-hrs activity diaries for three age groups: 3–5 year-olds, 6–8 year-olds, and 9–12 year-olds. According to the Kaiser Foundation [9], there is a gender impact on console video game playing, with boys spending an average of almost an hour a day playing (0:56) and girls just under fifteen minutes (0:14). One clear reason for the disparity in this age group is that girls lose interest in computer games as they enter their teenage years, whereas boys do not. Other studies have also reported gender differences in children's video game playing [10] [11], and the 2003 consumer survey by Interactive Digital Software Association [12] showed that approximately 72% of the most frequent players are boys or men. In arcades, video games are more often played by boys or men than by girls or women [13]; meanwhile, according to [14], approximately equal numbers of men and women in college play video games.

Gender is also a relevant issue regarding video games' content. More video games are male-oriented than female-oriented where Scharrer [15] examined 1,054 video game advertisements from video game magazines and reported that the ratio of male to female characters was greater than three to one. After their analysis of 597 characters from 47 randomly selected Nintendo 64 and PlayStation games, Beasley and Standley [16] found that only approximately 14% of those characters were women. In addition, female characters were depicted with more exposed skin than male characters. In a study of video game reviews on an Internet site, Ivory [17] similarly reported under-representation and sexualization of female video game characters. In none of these previous studies, though, were players' perspectives of gender-related content assessed.

Differential video game-playing by men and women may also be related to other aspects of individuals' lives. In [18], the author shows that brief exposure of children to a prosocial video game (PVG) increased their prosocial thoughts and prosocial behaviors. More precisely, boys reported higher accessibility of prosocial thoughts and more prosocial

behaviors than girls. The PVG effect on prosocial behaviors was mediated by prosocial thoughts. These findings suggest that increasing PVG exposure and training prosocial thoughts were effective ways to promote the positive development of prosocial behavior during early childhood. The research in [19] suggested that players' in-game motivational experiences can contribute to affective well-being, but they do not affect the degree to which play time relates to well-being. On other hand, the "displacement hypothesis" has been used previously to explain how children's television viewing may affect their other activities [20]; time spent in one activity displaces time that could have been used to do something else. Similarly, Gentile, Lynch, Linder, and Walsh [21] reported that both the amount of game playing and exposure to game violence were negatively linked to poor school grades in eighth- and ninth-grade students. Sixth- and ninth-grade boys were more likely than girls of the same age to indicate that they spent their free time playing computer/video games, and they were more likely to select "none" when asked how much time per day they spent reading for pleasure [22]. Video game playing has been linked with hostility [21], which could negatively impact interpersonal relationships in a variety of ways. Also, if men spend more time than women playing video games, romantic relationships may be negatively impacted because "couple time" could be displaced by time spent gaming, which could result in interpersonal conflict.

### III. METHODOLOGY

The study began by forming focus groups, which are effective in exploring and examining participants' perspectives and concerns by allowing them to create new questions and concepts [23]. Focus group sessions are social gatherings, usually of six to eight participants [24]. Morgan [25] reported that it can take up to 32 telephone calls or personal visits to recruit just eight participants for a group. Participants are encouraged to debate the relevant issues and to develop their opinions and thoughts, as they would in real-world situations [26]. A moderator presents the focus of the discussion and helps to elicit conflicting arguments without judging the participants' opinions [27]. The approach used to analyze the resulting data was to categorize quotations from the focus groups into types of description, i.e., concepts, and then to compare them with targeted concepts of privacy perspectives [28]. Four focus groups were formed during March 2023 with 26 participants in total: a men-only group (seven participants), a women-only group (six participants), and two groups of mixed-gender (four women and three men, and three women and three men). All the focus group participants were college students aged 18–25 years and a snowballing technique was used to recruit them. Prior to each focus group, participants' consent obtained for ethical considerations. Conducting the four focus group discussions early in the study was useful, as they provided preliminary information on which to focus on the next step of the data collection.

A questionnaire was then designed to evaluate the possible influence that gender had on college students' perspectives and participation in video games. The questionnaire was based on the outcomes of the focus groups and face-validated using exploratory interviews; some items were rephrased in order to reflect the intended meaning, while others were deleted or

added. To define the final list of statements, respondents were asked to identify whether the proposed items from the questionnaire represented their perspectives toward video games and to indicate some additional items that they considered important for investigation. The questionnaire consisted of two parts: the first gathered gender and age information, and the second included 54 five-point Likert statements ranging from "strongly disagree" to "strongly agree." The questionnaire was created using Google Forms and the random target subjects of the questionnaire were college students in the KSA aged 18–25 years. The questionnaire took 10–15 minutes to complete. After collecting the responses, the data were entered into the computer and processed using the Statistical Package for the Social Sciences (SPSS V.20), which is a widely used program for statistical analysis in the social sciences.

## IV. RESULTS

### A. Focus Group Results

Analysis of the data from the focus groups revealed four main areas related directly to the perspectives and participation of students in video games in the KSA: playing practices; bullying; the pros and cons of video games; and the most used consoles and most popular games.

Playing practices consisted of variables that relate to activities such as spending money on video games, watching professionals playing, gender issues in video games, and the effect of playing on social and academic life. Bullying in video games comprised questions related to someone being bullied and practices to avoid bullying. The third area referred to possible advantages and disadvantages of playing video games, such as isolation or engagement with others, wasting money or time, health and religious effects, and improvements in mental agility and concentration. The final area includes variables relating to the most used consoles and the most popular video games.

### B. Questionnaire Results

There were 2,756 responses to the questionnaire, of which 39.2% were from men and 60.8% were from women. However, 469 responses were excluded from the results because they were from respondents over 25 years of age; therefore, 2,287 responses were included in the study.

1) *Playing time:* Table I shows that when participants were asked about how much time they spent playing during the previous week, 26.9% did not play, and more women did not play than men. It can also be seen that 23% of participants played 1 to 2 hrs per week, and again most of them were women. However, the percentage of male players was higher for the last three categories: 2 to hrs, 6 to 10 hrs, and every day. In addition, the chi-squared test was used to determine the gender differences and playing hours in video games, and it was found that the p - value showed there was a statistically significant difference between men and women, and the number of hours of play differed according to gender. From the table below it can be seen that men prefer to spend more hours on video games than women.

TABLE I. TIME SPENT PLAYING DURING THE PREVIOUS WEEK

	Male	Female	Total (%)
None	132	610	742 (26.9%)
1–2 hrs	228	406	634 (23%)
2–5 hrs	226	182	408 (14.8%)
6–10 hrs	122	72	194 (7%)
Every day	191	118	309 (11.2%)
<b>Total (%)</b>	899 (39.3%)	1,388 (60.7%)	2,287(100%)

Chi-squared value = 21.41 *p* - value = 0.000\*

\**p* - value is statistically significant at 0.05

Respondents who did not play video games in the last week were presented with a choice of reasons why. The mean, standard deviation, and level of agreement of the responses are illustrated in Table II. The reasons chosen by respondents reflect a generally positive perspective, as the reasons for not playing were not lack of money or lack of ability to play, but rather their lack of time or enthusiasm to play.

TABLE II. REASONS FOR NOT PLAYING IN THE PREVIOUS WEEK

Reason	Mean	Standard Deviation	Level of Agreement
I don't have time	3.86	1.30	High
I don't like video games	3.90	1.29	High
I don't have enough money to play	2.39	1.37	Low
I am not good at video games	3.14	1.41	Medium

2) *Playing practices*: Table III demonstrates the respondents' answers to their playing practices, along with the mean, standard deviation, and level of agreement where the average mean was 3.28, which is medium.

The respondents' perspectives for all items varied between high and medium, where the highest mean of 3.72 was for item 1 (“For fun, I prefer to watch videos of other people playing”) followed by item 2 (“I will not pay for a video game that I never tried before; I have to try it first”) with a mean of 3.65, which indicates respondents' mindfulness of spending money on games. Items 12 and 13 have the lowest means of 2.93 and 2.70, respectively, indicating that respondents believe that video games do not have a significant effect on their academic performance or their social relationships, or that they are not sure.

3) *Bullying in video games*: Tabel IV presents the mean, standard deviation, and level of agreement regarding 13 items relating to bullying in video games. The mean was 3.46 in the high level, which indicates that respondents have high awareness regarding this topic, and as noted the mean for all the 13 items related to this topic ranged between 2.90–3.89, indicating a high to medium level. The highest mean, 3.89, is for item 1 (“I see a lot of bullying in video games”), followed by item 2 (“I play only with my friends”) where the mean was 3.84. The lowest means, 3.08 and 2.90, were for items 12 and 13, respectively (“I don't care if I witnessed someone being bullied,” and “Bullies are always female characters”); the level of perspectives was medium, indicating that respondents' are neutral about these two items.

TABLE III. PLAYING PRACTICES

	Item	Mean	Standard Deviation	Level of Agreement
1	For fun, I prefer to watch videos of other people playing	3.72	1.26	High
2	I will not pay for a video game that I never tried before; I have to try it first	3.65	1.28	High
3	I watch videos of other people playing just to learn from them	3.57	1.25	High
4	Most famous video games are masculine	3.48	1.14	High
5	In video games female characters are always sexually provocative	3.40	1.28	Medium
6	Playing affects my sleeping hours negatively	3.35	1.26	Medium
7	I might pay for a video game that I never tried before	3.24	1.38	Medium
8	I don't like to deal with people who play video games a lot	3.20	1.24	Medium
9	There are no famous video games that are feminine	3.20	1.19	Medium
10	I might pay to update a game or buy a weapon	3.13	1.39	Medium
11	I always pick a male character	3.04	1.40	Medium
12	Playing affects my academic performance negatively	2.93	1.29	Medium
13	Playing affects my relationships with others negatively	2.70	1.31	Medium
	Total	3.28	0.62	Medium

TABLE IV. BULLYING IN VIDEO GAMES

	Items	Mean	Standard Deviation	Level of Agreement
1	I see a lot of bullying in video games	3.89	1.05	High
2	I play only with my friends	3.84	1.14	High
3	Usually, if the team is all female players, they would refuse a guy playing with them	3.75	1.10	High
4	I do play with strangers picked randomly by the game.	3.71	1.08	High
5	Male characters are more professional in video games	3.69	1.23	High
6	Bullies are always men	3.57	1.14	High
7	If I witness someone being bullied, I always stand out	3.45	1.19	High
8	I do play with strangers from the opposite sex picked randomly by the game	3.41	1.25	High
9	Bullying is always against female characters	3.37	1.20	Medium
10	Women are more professional in video games	3.15	1.19	Medium
11	Usually, if the team is all male players, they would refuse a female player with them	3.11	1.31	Medium
12	I don't care if I witness someone being bullied	3.08	1.30	Medium
13	Bullies are always female characters	2.90	1.14	Medium
	Total	3.46	0.63	High

4) *Pros and cons of playing video games:* Participants in the focus groups were asked about the pros and cons of playing video games. Table V shows that the average mean was 3.88, with a high level, and this falls within the category of “agreed,” which indicates the respondents' agreement towards the pros and cons of playing video games and that they have awareness of the listed items. It is also noted that the mean of their perspectives ranged between 3.56–4.31 for the items listed and the highest mean was 4.31 for item 1 (“Spending fun time’ is an advantage of playing”). The lowest, 3.57 and 3.56, were for items 9 and 10, respectively (“Neglecting work/study’ is a disadvantage of playing,” and “Wasting money’ is a disadvantage of playing”) The perspective level was high which indicates that the respondents' approval of these cons is high, confirming that they concur on the defects of video games including their impact on work, academic achievement, and financial loss.

TABLE V. PROS AND CONS OF PLAYING VIDEO GAMES

	Items	Mean	Standard Deviation	Level of Agreement
1	Spending fun time’ is an advantage of playing	4.31	0.78	Very High
2	Living exciting moments in artificial reality’ is an advantage of playing	4.18	0.89	High
3	Evolving player's smartness and reflections’ is an advantage of playing	4.07	0.97	High
4	Affecting player's sight negatively or causing obesity’ is a disadvantage of playing	3.88	1.09	High
5	Knowing other players and get along with them’ is an advantage of playing	3.86	1.08	High
6	Neglecting prayers’ is a disadvantage of playing	3.84	1.20	High
7	Isolating players from community’ is a disadvantage of playing	3.80	1.09	High
8	Wasting time’ is a disadvantage of playing	3.74	1.17	High
9	Neglecting work/study’ is a disadvantage of playing	3.57	1.18	High
10	Wasting money’ is a disadvantage of playing	3.56	1.18	High
	Total	3.88	0.65	High

5) *Impact of gender:* In addition to the mean and standard deviations, the independent samples test was used to identify gender impact on respondents' perspectives towards playing practices, bullying, and the pros and cons of playing video games. Table IV shows that the p - value is statistically significant in both playing practices and the pros and cons of video games. This indicates that the perspectives of male respondents differ from those of female respondents towards these two areas, where the mean for men’s perspectives toward playing practices is higher than the women’s, indicating that men have a better perception of their playing

practices; however, it is the opposite case in the second area, where women’s perspectives toward pros and cons for playing video games is higher, indicating that they have a better perception of this area. However, it should be noted that there are no differences in male and female respondents' perspectives towards bullying in video games, meaning that their opinions are similar in this area. Fig. 1 below demonstrated the mean and standard deviations for males and females participants towards the three areas.

TABLE VI. IMPACT OF GENDER

Item	Gender	Sample Size	Mean	Standard Deviation	p - value
Playing practices	Male	767	3.34	0.59	0.000*
	Female	778	3.22	0.64	
Bullying	Male	767	3.46	0.62	0.911
	Female	778	3.45	0.64	
Pros and cons	Male	767	3.83	0.62	0.005*
	Female	778	3.93	0.67	

\*p - value is statistically significant at 0.05

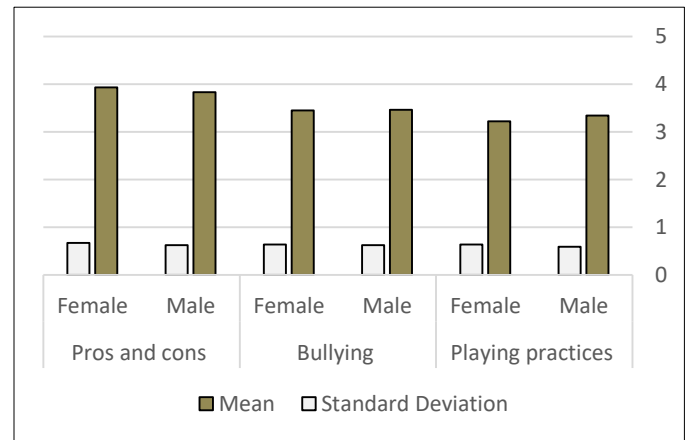


Fig. 1. Gender impact.

6) *Most used game console and most popular game:* The independent samples test was used in addition to mean and standard deviations to identify the influence of gender on the game consoles most used by male and female respondents, as presented in Table VII. The p - values in the table below indicate a statistical significance in the following consoles: Sony PlayStation, mobile phone, PC, and Xbox; this indicates that men's opinions differ from women’s towards these consoles, and it is clear from the table that the differences were in favor of males respondents, indicating that men have a higher preference than females for their use. Meanwhile, the p - value in the table is not statistically significant in the perspectives of male and female respondents towards Wii and Nintendo, indicating that the perspectives of male respondents do not differ from those of women towards these consoles. Fig. 2 demonstrated the mean and standard deviations for males and females participants towards the different game consoles.

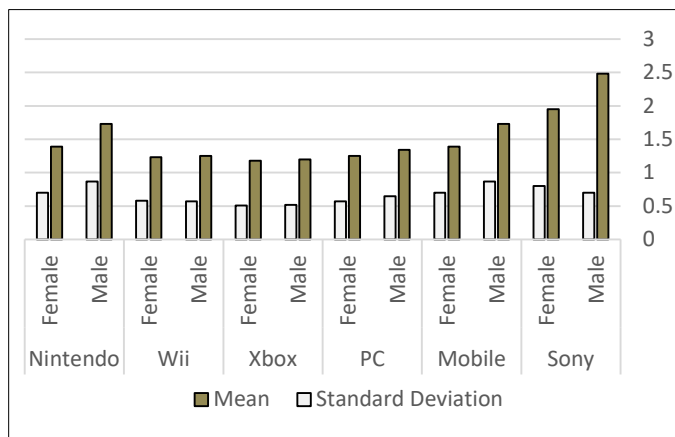


Fig. 2. Game consoles.

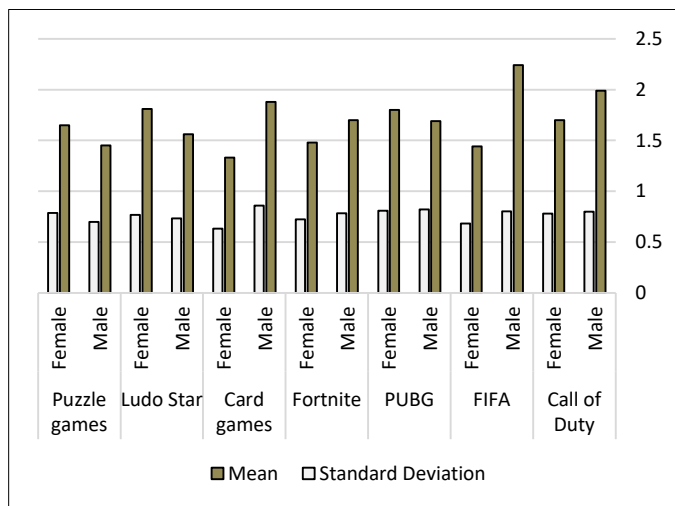


Fig. 3. Popular games.

Meanwhile, seven video games were mentioned by participants in the focus groups: Call of Duty, FIFA, PUBG, Fortnite, card games, Ludo Star, and puzzle games. Call of Duty is a video game series starting in 2003 and originally focused on the World War II setting, FIFA is a discontinued football video game, PUBG (previously known as PlayerUnknown's Battlegrounds) is a battle royal game, Fortnite is an online fighting video game released in 2017, card games are any game using playing cards as the primary device with which the game is played such as Solitaire, Spider Solitaire, Hearts, etc. Ludo Star is an online board dice video game, puzzle games are make up a broad genre of video games that emphasize puzzle solving. The types of puzzles can test problem-solving skills, including logic, pattern recognition, sequence solving, spatial recognition, and word completion. By exploring gender influence on the popularity of these games, Table VIII shows that the  $p$  - value is statistically significant in the perspectives of male and female respondents towards all the games, and this indicates that the perspectives of men differ from those of women towards playing those games. The differences were in favor of male respondents for Call of Duty, FIFA, Fortnite, and card games, and this indicates that men have a better perception than women of these games or that these games suit men better than women. On the other hand,

the differences were in favor of women for PUBG, Ludo Star, and puzzle games, and this indicates that women have a better perception than men of these games or that these games are more suitable for women than men. Fig. 3 demonstrated the mean and standard deviations for males and females participants towards popular games.

TABLE VII. MOST USED CONSOLES

Console	Gender	N	Mean	Standard Deviation	$p$ - value
Sony PlayStation	Male	767	2.48	0.70	0.000*
	Female	778	1.95	0.80	
Mobile phone	Male	767	1.73	0.87	0.000*
	Female	778	1.39	0.70	
PC	Male	767	1.34	0.65	0.000*
	Female	778	1.25	0.57	
Xbox	Male	767	1.20	0.52	0.003*
	Female	778	1.18	0.51	
Wii	Male	767	1.25	0.57	0.367
	Female	778	1.23	0.58	
Nintendo	Male	767	1.73	0.87	0.383
	Female	778	1.39	0.70	

\* $p$  - value is statistically significant at 0.05

TABLE VIII. MOST POPULAR VIDEO GAME

Video game	Gender	N	Mean	Standard Deviation	$p$ - value
Call of Duty	Male	767	1.99	0.798	0.000*
	Female	778	1.70	0.780	
FIFA	Male	767	2.24	0.801	0.000*
	Female	778	1.44	0.684	
PUBG	Male	767	1.69	0.822	0.000*
	Female	778	1.80	0.810	
Fortnite	Male	767	1.70	0.782	0.003*
	Female	778	1.48	0.723	
Card games	Male	767	1.88	0.859	0.000*
	Female	778	1.33	0.631	
Ludo Star	Male	767	1.56	0.732	0.000*
	Female	778	1.81	0.769	
Puzzle games	Male	767	1.45	0.700	0.000*
	Female	778	1.65	0.787	

\* $p$  - value is statistically significant at 0.05

## V. DISCUSSION

Regarding the first research question about the main areas that identify the practices and perceptions of video game players in the KSA, the analysis found four main areas: playing practices, bullying, pros and cons of video games, and most used console and most popular game.

Meanwhile to answer the second research question: Are there differences between male and female players' perceptions

towards video games in the KSA? The similarities and differences between male and female players' perceptions towards video games have been highlighted. Regarding the playing practices, the study found that approximately one-third of the participants did not play at all in the week before the questionnaire, and the majority of these were women. For participants who played video games, the study found that more men than women prefer to play, and for longer hours, which supports the original hypothesis. This is supported by [8], [9] and [29], which reported that the percentages of daily and weekly male players are higher than those of female players. In addition, men's perspectives toward playing practices mentioned in Table III are higher than women's, which indicates that men have a better perception of their playing practices and wide agreement regarding the items mentioned in the table, particularly the first four items: watching others playing for fun or to learn something, not buying a game without trying it first, and the masculinity of famous games. Also, gender difference is noticeable in the responses of participants regarding the pros and cons of playing video games where, in this case, women have a better perception of them and wide agreement regarding the items mentioned in Table V such as playing for fun, increasing thinking and response abilities, and getting friendly with strangers are the advantages of playing video games, while the disadvantages are the negative impact on academic/work performance and social relationships, and wasting time and money. According to [29], 17.7% of players in the KSA use smartphones to play, 17.6% of them use Sony PlayStations, and the rest of the consoles are used by less than 5% of players. This study found the opposite, where the Sony PlayStation is used more than smartphones, and men's usage is higher than women's. Moreover, [29] specify the order of the most popular games in playing consoles as follows: FIFA, PUBG, Fortnite, Roblox, and Call of Duty. Similarly, the most popular games on smartphones are as follows: Subway Surfers, Snake.io, Roblox, Ludo Star, Yalla Ludo, and PUBG MOBILE. This order of popular games did not consider the gender aspect. In this study, the order of popular games would be the following: FIFA, Call of Duty, PUBG, Ludo Star, Fortnite, and Card games at the same level, and then finally puzzle games. The gender impact appears in men favoring the first two along with Fortnite and card games, while women favor PUBG, Ludo Star, and puzzle games. On the other hand, gender differences did not occur in one area, namely bullying in video games. This indicates that male and female players' opinions are similar, and there is wide agreement about the items in this area. These are mentioned in Table IV particularly the first eight items about the frequency of bullying and specifically male bullies in video games and the general preference of only playing with friends if possible; if not, it is still acceptable to team up with strangers picked out randomly by the game even if they are from the opposite sex.

The third research question regarding the interference of video games with academic preparation and interpersonal relationships was answered by the analysis of the data in V, where participants highly agreed that getting friendly with strangers are one of the advantages of playing video games, while the disadvantages are the negative impact on academic/work performance and social relationships.

The current study was exploratory and the self-reported approach used were as transparent as possible to enable others to examine and build upon this work, however, moving beyond this initial exploration of players' practices and perceptions to a more confirmatory approach, further research should follow different approach relying on collecting practical practices. Following this will result in a more reliable knowledge base for game developers.

## VI. CONCLUSIONS

In the present study, the authors explored gender differences in video game playing among college students in the KSA, covering players' perspectives of the male and female characters' roles in the video games that they played, and the potential interference of gaming with other aspects of college students' lives. It found clear evidence supporting the high impact of gender on the participation and perceptions of college students in the KSA, as predicted. In addition, the four main areas identified can serve as the foundation for the development of appropriate game-marketing strategies for game vendors in the KSA. Game developers who are targeting the KSA game market can use the results of this study as the basis for developing video games by considering the impact of gender and the practices and perceptions of Saudi players detailed in this study. However, selecting college students puts a limitation on how generalizable our results are. Moreover, collecting data from participants using self-reports instead of objective measures of technology use can be other limitation to the current study. The findings in this study advance the knowledge of this subject and future research could be expanded by increasing the number of questions asked regarding students' academic performance and income or expanding the study range to include players other than college students. In addition, as this study was exploratory, we recommend other researchers to examine and build upon our work and collecting data regarding players' practices by using of more confirmatory approach.

## REFERENCES

- [1] International Trade Administration U.S. Department of Commerce, 06 07 2022. [Online]. Available: <https://www.trade.gov/country-commercial-guides/saudi-arabia-travel-tourism-and-entertainment> [Accessed 05 03 2023].
- [2] Newzoo, 29 01 2021. [Online]. Available: <https://newzoo.com/insights/articles/playing-and-spending-habits-in-saudi-arabias-games-market/>. [Accessed 05 03 2023].
- [3] C. A. Anderson and B. J. Bushman, "Effects of violent video games on aggressive behavior, aggressive cognition, aggressive affect, physiological arousal, and prosocial behavior: A metaanalytic review of the scientific literature," *Psychological Science*, vol. 12, no. <https://doi.org/10.1111/1467-9280.00366>, p. 353–359, 2001.
- [4] K. Haninger and K. M. Thompson, "Content and ratings of teen-rated video games.," *Journal of the American Medical Association*, vol. 291, no. [doi:10.1001/jama.291.7.856](https://doi.org/10.1001/jama.291.7.856), p. 856–865, 2004.
- [5] W. G. Kronenberger, V. P. Mathews, D. W. Dunn, Y. Wang, E. A. Wood and J. J. Larsen, "Media violence exposure in aggressive and control adolescents: Differences in self- and parent-reported exposure to violence on television and in video games.," *Aggressive Behavior*, vol. 31, no. <https://doi.org/10.1002/ab.20021>, p. 201–216, 2005.
- [6] Y. J. Halbrook, A. T. O'Donnell and R. M. Msetfi, "When and How Video Games Can Be Good: A Review of the Positive Effects of Video Games on Well-Being," *Perspectives on Psychological Science*, 14(6), p. 1096–1104. <https://doi.org/10.1177/17456916198638>, 2019.

- [7] C. Comeras-Chueca, L. Villalba-Heredia, M. Pérez-Llera, G. Lozano-Berges, J. Marín-Puyalto, G. Vicente-Rodríguez, A. Matute-Llorente, J. Casajús and A. González-Agüero, "Assessment of Active Video Games' Energy Expenditure in Children with Overweight and Obesity and Differences by Gender," *International Journal of Environmental Research and Public Health*, vol. 17(18), p. 6714. <https://doi.org/10.3390/ijerph17186714>, 2020.
- [8] J. C. Wright, A. C. Huston and E. A. Vandewater, "American children's use of electronic media in 1997: A national survey," *Journal of Applied Developmental Psychology*, vol. 22, pp. 31–47, [https://doi.org/10.1016/S0193-3973\(00\)00064-2](https://doi.org/10.1016/S0193-3973(00)00064-2), 2001.
- [9] V. Rideout, D. F. Roberts and U. G. Foehr, "Generation M: Media in the lives of 8–18 year-olds," Kaiser Family Foundation, Menlo Park, California, <https://files.eric.ed.gov/fulltext/ED527859.pdf>, 2010.
- [10] D. S. Bickham, E. A. Vandewater, A. C. Huston, J. H. Lee, A. G. Caplovitz and J. C. Wright, "Predictors of children's electronic media use: An examination of three ethnic groups," *Media Psychology*, vol. 5, pp. 107–137, [https://doi.org/10.1207/S1532785XMEP0502\\_1](https://doi.org/10.1207/S1532785XMEP0502_1), 2003.
- [11] E. H. Woodard and N. Gridina, "Media in the home 2000: The fifth annual survey of parents and children.," The Annenberg Public Policy Center., Philadelphia: University of Pennsylvania, [https://cdn.annenbergpublicpolicycenter.org/Downloads/Media\\_and\\_Developing\\_Child/mediasurvey/inhome.pdf](https://cdn.annenbergpublicpolicycenter.org/Downloads/Media_and_Developing_Child/mediasurvey/inhome.pdf), 2000.
- [12] Interactive Digital Software Association, "Demographic information.," 2005.
- [13] B. Jurica, K. Alanis and S. Ogletree, "Sex differences related to video arcade game behavior," *Psi Chi Journal of Undergraduate Research*, vol. 7, pp. 145–148, <https://doi.org/10.24839/1089-4136.JN7.3.145>, 2002.
- [14] R. Gardyn, "Got game?," *American Demographics*, vol. 25, no. 8, 2003.
- [15] E. Scharrer, "Virtual violence: Gender and aggression in video game advertisements," *Mass Communication & Society*, vol. 7, pp. 393–412, [https://doi.org/10.1207/s15327825mcs0704\\_2](https://doi.org/10.1207/s15327825mcs0704_2), 2004.
- [16] B. Beasley and T. C. Standley, "Shirts vs. skins: Clothing as an indicator of gender role stereotyping in video games," *Mass Communication & Society*, vol. 5, pp. 279–293, [https://doi.org/10.1207/S15327825MCS0503\\_3](https://doi.org/10.1207/S15327825MCS0503_3), 2002.
- [17] J. D. Ivory, "Still a man's game: Gender representation in online reviews of video games," *Mass Communication and Society*, vol. 9, pp. 103–114, [https://doi.org/10.1207/s15327825mcs0901\\_6](https://doi.org/10.1207/s15327825mcs0901_6), 2006.
- [18] H. Li and Q. Zhang, "Effects of Prosocial Video Games on Prosocial Thoughts and Prosocial Behaviors," *Social Science Computer Review*, 41(3), p. 1063–1080. <https://doi.org/10.1177/08944393211069599>, 2023.
- [19] N. Johannes, M. Vuorre and A. Przybylski, "Video game play is positively correlated with well-being," *Royal Society Open Science*, p. <https://doi.org/10.1098/rsos.202049>, 2021.
- [20] N. Shin, "Exploring pathways from television viewing to academic achievement in school age children," *Journal of Genetic Psychology*, vol. 165, pp. 367–381, <https://doi.org/10.3200/GNTP.165.4.367-382>, 2004.
- [21] D. A. Gentile, P. J. Lynch, J. R. Linder and D. A. Walsh, "The effects of violent video game habits on adolescent hostility, aggressive behaviors, and school performance," *Journal of Adolescence*, vol. 27, pp. 5–22, <https://doi.org/10.1016/j.adolescence.2003.10.002>, 2004.
- [22] M. A. Nippold, J. K. Duthie and J. Larsen, "Literacy as a leisure activity: Free-time preferences of older children and young adolescents," *Language, Speech, and Hearing Services in Schools*, vol. 36, pp. 93–102, [https://doi.org/10.1044/0161-1461\(2005/009\)](https://doi.org/10.1044/0161-1461(2005/009)), 2005.
- [23] J. Kitzinger and R. Barbour, *Developing Focus Group Research: Politics, Theory and Practice*, SAGE Publications, 1998.
- [24] M. Bloor, J. Frankland, M. Thomas and K. Robson, *Focus groups in social research*, London: SAGE, 2001.
- [25] D. L. Morgan, *Planning Focus Groups*, Thousand Oake, 1998.
- [26] P. Lunt and S. Livingstone, "Rethinking the focus group in media and communications-research.," *Journal of Communications*, vol. 46, no. 2, pp. 79–98, <https://doi.org/10.1111/j.1460-2466.1996.tb01475.x>, 1996.
- [27] C. Puchta and J. Potter, *Focus group practice*, London: SAGE, 2004.
- [28] J. Kitzinger, "Qualitative Research: Introducing focus groups," *BMJ*, vol. 311, no. 7000, pp. 299–302, <https://doi.org/10.1136/bmj.311.7000.299>, 1995.
- [29] Communications, Space & Technology Commission, CST, "Saudi Internet," *Communications, Space & Technology Commission*, <https://www.cst.gov.sa/en/indicators/saudiinternet/internt-saudi-2022.pdf>, 2022.

# Hyperparameter Tuning of Semi-Supervised Learning for Indonesian Text Annotation

Siti Khomsah<sup>1</sup>, Nur Heri Cahyana<sup>2</sup>, Agus Sasmito Aribowo<sup>3</sup>

Department of Data Science, Institut Teknologi Telkom Purwokerto, Indonesia<sup>1</sup>

Department of Informatics, Universitas Pembangunan Nasional Veteran Yogyakarta, Indonesia<sup>2,3</sup>

**Abstract**—A crucial issue in sentiment analysis primarily relies on the annotation task involving data labeling. This critical step is typically performed by linguists, as the nuanced meaning of text significantly influences its contextual interpretation. If there is a large volume of data, annotation is time-consuming and financially burdensome. Addressing these challenges, a semi-supervised learning annotation (SSL) that integrates human annotator and artificial intelligence algorithms emerges as a potent solution. Building accurate SSL needs to explore the best architecture, including a combination of machine learning and mechanism. This research aims to construct semi-supervised model annotation text by tuning the parameter of the machine learning algorithm to gain the most accurate model. This study employed a Support Vector Machine and a Random Forest algorithm to build semi-supervised annotation. Grid-Search and Random-Search were employed to tune the Random Forest and Support Vector Machine parameters. The semi-supervised annotation model was applied to annotate Indonesian texts. The outcomes signify that hyperparameter-tuning enhances SSL performance, surpassing the performance achieved using default parameters. The experiment also shows that the SSL annotation using a Support Vector Machine tuned by Grid Search and Random Search is more robust than the Random Forest algorithm. Hyperparameter tuning is also robust to training data that contains many manual labeling errors by experts.

**Keywords**—Text annotation; semi-supervised; parameter-tuning; grid search; random search

## I. INTRODUCTION

The challenge in text classification-based machine learning is data labeling. Each textual instance has meaning upon the context and grammatical nuance specific to each language. Thus, human intervention is vital for data labeling, as humans are adept at assessing the contextual relevance of text. Human annotator requires expertise in understanding the language context. However, labeling numerous volumes of data requires quite an amount of time and causes tiredness for the annotator - consequently, the objectivity loss of the labeling process. Hence, automation annotation is highly needed before the dataset feeds into a machine learning classifier.

There has been a new development of an annotator-based machine built by learning knowledge of the language experts or humans. SSL uses a small sample of data annotated by language experts and then uses the sample to build a training model. Then, the training model is used for the annotation of unlabeled data. The SSL model draws expert insight from a small sample to build robust annotator unlabeled data. Here, the SSL challenge is to produce a reliable and precise model.

In previous related research, a semi-supervised learning (SSL) model was developed to classify text [1-4]. The performance of semi-supervised text annotation still needs to improve. Al-Laith et al. used LSTM and FastText for annotating Arabic text with only three classes, resulting in an accuracy of 69.4% on the SemEval 2017 dataset. In contrast, the best system achieved a 63.38% F1 score, while on the ASTD dataset, performance improved from 64.10% to 68.1% [4]. Aydin and Güngör [5] combined semi-supervised and unsupervised methods and implemented the J-48 tree, Support Vector Machine, and Naive Bayes algorithms. As a result, the accuracy is more than 90%. However, this model has not proven its performance for multi-class datasets and other language datasets. Alahmary and Al-Dossari [2] applied Naive Bayes as a semi-supervised learning classifier with high accuracy (83%). However, researchers have yet to test this model on other datasets, and we do not know its performance when utilized for SSL in other languages. So, developing a semi-supervised text annotator for Indonesian text that utilizes multiple machine learning models and various vectorizers is expected to yield improved performance. Previous related Research SSL in Indonesian by NurHeri Cahyana et al. annotates hate speech [3]. The maximum accuracy in those research employing KNN and TF-IDF is only 59.68%. SSL model in [3] has a weakness: the model has not been applied to other Indonesian text datasets and has not tried other model combinations.

Many ways are done to gain a high-performance model, such as identifying data samples with hesitant labels and removing training data with unreliable labels. Those can improve the quality of the training data and improve classifier performance [6]. Another approach to improve model performance is to overcome by trying various amounts of training data proportions when building the annotation model, then applying the ensemble method, using several machine learning to classify the same data consensus on classification decisions using confidence values [7]. Performance machine learning depends on parameter setting [8]. Some ways to leverage the performance of machine learning are tuning the parameters [9–11] and applying optimization [12-14]. The goal of tuning these parameters is to find the best combination parameter [15]. Tuning parameters can significantly impact the model performance and finding the right combination involving experimentation and iterative adjustments [16]. A tuning parameter is a parameter that is not learned directly from the data during the training process of a machine-learning algorithm. Instead, it is a value set before training to control the algorithm behavior. Unlike the default individual parameters of



a model, which are learned from the training data, tuning parameters are set by the user before training. Fine-tuning parameters can have a substantial impact on the model performance [17].

Besides tuning individual algorithm parameters, hyperparameter techniques such as Grid Search and Random Search can help find the best parameter combination. Several researchers in text labeling or other domains employ Grid Search [9], [18], [19] and Random Search [10], [11], [20], [21] to enhance the accuracy and performance of machine learning algorithm. Not all parameter tuning techniques are in tune with machine learning algorithms and the data they handle. In previous research, several algorithms that can generally be tuned include Support Vector Machine (SVM), Random Forest (RF), Logistic Regression, Naive Bayes, Neural Networks, and Gradient Boosting. It is necessary to research the best hyperparameters for each algorithm. This research aims to leverage the performance of semi-supervised text annotation through tuning parameters Support Vector Machine and Random Forest using Grid Search and Random Search. Different language corpus needs an appropriate architecture model for automated text labeling. This research uses the Indonesian dataset for the research material. Thus, SSL architecture for the Indonesian dataset becomes the focus.

Our research proposed SSL with the SVM and Random Forest as classifier. Random Search and Grid Search as hyperparameter tuning to obtain the most optimal model. The research problem boundary employs hyperparameter tuning to gain high-performing SSL models for Indonesian text annotation—the performance metric using F1-Score and accuracy. This research is limited to utilizing two machine learning algorithms, Random Forests and Support Vector Machine, and two hyperparameters technique Grid Search and Random Search. We divide the article into four sections: introduction, methods, results and discussion, and conclusions.

## II. Methods

The following section describes the research steps, including data collection, data cleaning, feature extraction, building classifier model annotation, and evaluation.

### A. Data Collection

Due to several reasons, this research has used various datasets to test the Semi-Supervised Text Annotation Model. The primary rationale behind this approach is to enhance the model's ability to generalize and effectively accommodate variations within the data. By subjecting the model to evaluation using diverse datasets, the model can objectively assess its performance across different domains. Also, this approach can examine the model's efficacy in managing linguistic heterogeneity. Additionally, testing on various datasets has comprehensively evaluated the model performance. Thus, this research endeavor involves the utilization of three publicly available datasets, as described in Table I, for the explicit purpose of rigorous testing and analysis.

TABLE I. DATASETS USED FOR MODEL ASSESSMENT

No	Dataset	Instance	Source
1.	Hate Speech	13168	<a href="https://github.com/okkyibrohim/id-multi-label-hate-speech-and-abusive-language-detection/blob/master/re_dataset.csv">https://github.com/okkyibrohim/id-multi-label-hate-speech-and-abusive-language-detection/blob/master/re_dataset.csv</a>
2	Sentiment Ridife	10805	<a href="https://github.com/ridife/dataset-idsa/blob/master/Indonesian%20Sentiment%20Twitter%20Dataset%20Labeled.csv">https://github.com/ridife/dataset-idsa/blob/master/Indonesian%20Sentiment%20Twitter%20Dataset%20Labeled.csv</a>
3	IndoNLU Sentiment	12759	<a href="https://github.com/IndoNLP/IndoNLU/tree/master/dataset/smsa_doc-sentiment-prosa">https://github.com/IndoNLP/IndoNLU/tree/master/dataset/smsa_doc-sentiment-prosa</a>

Before their utilization in the testing phase, the four datasets were processed by the same steps, including data cleaning, word embedding, and feature extraction. The subsequent section elaborates on the empirical outcomes of analyzing these four distinct datasets.

### B. Data Cleaning and Preprocessing

The datasets used in this experiment comprised comments in the Indonesian language. The cleaning purpose is to clear up the dataset from noise such as punctuation, numerical character, and stop words. The clean data was transformed into a vector through several stages, including tokenization (unigram, bigram, and trigram), stemming, and finally, turning the stem into the vector using TF-IDF. Tokenizing onto unigram is breaking down a piece of a sentence into individual units. Bigram is a pair of sequence words within a sentence, while trigram refers to three sequence words within a sentence. Unigram, bigram, and trigram are N-gram types in which  $N$  is any number. N-Gram with  $N$  is two or more, often used to capture more contextual information than a single word.

### C. Word Embedding and Feature Extraction

The Bag of Words (BoW), often mentioned as Term-Frequency (TF), constitutes an algorithm employed to determine the weight of individual words within a document. The weight of Term-Frequency is computed by quantifying the occurrence of the term  $t$  in document  $D$  and dividing it by the total count of words present in document  $D$ . The underlying objective is to identify unique words that can serve as an essential document feature. The large documents generate a big matrix. Given a number feature of  $N$  and a sum document of  $D$ , the feature matrix has dimensions of  $N \times D$ . TF is the occurrence of a word  $t$  in document  $D$ , computed as in Eq. (1).

$$TF_{t,d} = \frac{n_{t,d}}{\text{Total number of terms in document}} \quad (1)$$

The  $TF_{t,d}$  is the frequency of term  $t$  in document  $d$ , where  $t$  is a term (word within a sentence),  $f$  is the number of occurrences of term  $t$  in document  $d$ , and  $n_{t,d}$  is the number of terms  $t$  in document  $d$ .

TF-IDF (Term Frequency-Inverse Document Frequency) is the weight computed by multiplication between TF and IDF. IDF (Inverse Document Frequency) is a value that indicates how important a word is in the entire document in the dataset. A word with a lower TF-IDF value is considered less important, and vice versa. Words that appear frequently throughout the document are considered as less important words. The weight IDF of a document calculated as in Eq. (2)

$$IDF_d = \log\left(\frac{\text{Number of Document}}{\text{Number of document with term } t_i}\right) \quad (2)$$

To compute the TF-IDF (Term Frequency-Inverse Document Frequency), the TF value is multiplied by the IDF value according to the following formula, as in Eq. (3)

$$TFIDF_{t,d} = tf_{t,d} \times idf_d \quad (3)$$

#### D. Model of Semi-Supervised Text Annotation

Random Forest is a robust ensemble learning algorithm widely used in machine learning for classification. Random Forest defines the target class by combining multiple decision tree outputs to yield a single outcome. As suggested by its name, a "forest" comprises numerous trees generated through bagging or bootstrap aggregating. Each tree in the Random Forest produces class predictions, with the majority class prediction as the candidate prediction model. Increasing the number of trees leads to enhanced accuracy and mitigates overfitting concerns. Random Forest is known for its robustness and resistance to overfitting. Tuning its parameters can improve performance on specific datasets.

SVM is a robust machine learning algorithm widely used for classification tasks. It is particularly effective for tasks involving complex data distribution and when clear dividing lines are needed to differentiate between classes. SVM aims to find the optimal hyperplane that best separates different classes of data points in a high-dimensional space. The basic idea of SVM is to find a hyperplane that maximizes the margin between classes of data points. The margin is defined as the distance between the hyperplane and the nearest data points from each class. The idea is to choose the hyperplane with the largest margin, which is expected to classify well to new data with no class yet. However, SVM's effectiveness depends on properly tuning parameters to make the data separable.

#### E. Proposed Semi-Supervised Learning Architecture

This research proposes a novel architecture for the semi-supervised learning (SSL) model with tuning parameters, depicted in Fig. 1. The SSL workflow initiates by utilizing an annotated dataset encompassing training, testing, and unlabeled data, as in Fig. 1. Training and testing data are manually labeled by an Indonesian language expert.

The process begins with word embedding, transforming textual data from the training set into vectors using the TF-IDF technique. This word embedding procedure generates three distinctive vectors: unigram, bigram, and trigram. These three vector representations subsequently serve as inputs for building three separate models employing Random Forest and Support Vector Machine (SVM). The algorithm Random Forest and SVM parameters were tuned to gain the best classification model.

Following the stacking principle, these three distinct models operate independently. Each model participates in the annotation of the unlabeled data. Then, the unlabeled data is annotated by each model, resulting in pseudo-labels of three sets of datasets, each classified according to one of the three models. A pseudo-label has high confidence if the cumulative weight assigned to it divided by the cumulative weight of all models is high. This confidence value is compared to a

predefined threshold. This threshold serves as a criterion to identify annotated data (with pseudo-labels) with confidence values deserving inclusion as part of the training data. Then, documents with high-ranking confidence and existing training data are mixed. Through this innovative approach, the SSL process harnesses the strengths of multiple models to improve predictions on unlabeled data iteratively. Considering confidence values and applying the threshold optimizes integrating pseudo-labeled data into the training set, ultimately enhancing the model's performance.

Fig. 2 describes the main algorithm of SSL. The process started by reading the labeled training dataset (DT), testing dataset (DTest), and unlabeled dataset (UN). All three datasets are transformed into the vector with feature unigrams, bigrams, and trigrams. Then, the model-building process is shown in lines 9-11, the hyperparameter process is in lines 12-14, the annotation process with the best parameter model is in lines 15-17, and the line 18-21 is the voting process.

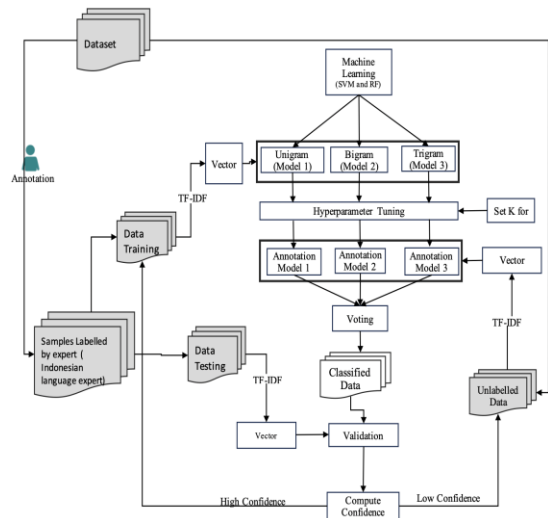


Fig. 1. SSL text annotation using parameter tuning.

```

1  DEF SSL(X,ML,HyP):
2  READ DT // Data Training(X,y)
3  READ DTest // Data Testing(X,y)
4  READ UN // Unlabeled Data(X)
5  VTestUni, VTestBi, VTestTri = TFIDF (DTest, ngram=1,2,3)
6  Loop three times or until no unlabelled dataset :
7      VTrainUni, VTrainBi, VTrainTri = TFIDF(DT, ngram=1,2,3)
8      VUnlabUni, VUnlabBi, VUnlabTri = TFIDF(UN, ngram=1,2,3)
9      Mod1 = ML.Train(VTrainUni)
10     Mod2 = ML.Train(VTrainBi)
11     Mod3 = ML.Train(VTrainTri)
12     Accuracy1,bestparam1=Hyperparam(Mod1,HyP)
13     Accuracy2,bestparam2=Hyperparam(Mod2,HyP)
14     Accuracy3,bestparam3=Hyperparam(Mod3,HyP)
15     Label[1]=Mod1.PredictProba(VUnlabUni, bestparam1)
16     Label[2]=Mod2.PredictProba(VUnlabBi, bestparam2)
17     Label[3]=Mod3.PredictProba(VUnlabTri, bestparam3)
18     For J = 1 to LEN(UN):
19         Newlabel[J], WeightLabel[J]=Voting(Label[1,2,3].RecNo[J]
20         Move(UN[J],Newlabel[J]) to DT
21 Output(DT)
22 Validate(DT) with Accuracy, F1Score
23 END
24
25 BEGIN
26 ML=['RF', 'SVM']
27 Hyperparameter=['GridSearch', 'RandomSearch']
28 For X in ML:
29     DO SSL(X, ML,Hyperparameter)
28 END
    
```

Fig. 2. Algorithm of proposed SSL.

### F. Tuning Techniques

Grid Search and Random Search were applied for tuning both SVM and RF. *Grid Search* is a hyperparameter tuning technique that systematically searches for a machine learning optimal combination model of hyperparameter values. It involves defining a grid of possible hyperparameter values and then evaluating the model performance using each combination of these values through cross-validation. The best parameter combination selected is the optimal set of hyperparameters [18].

Random Search does not explore all combinations like Grid Search. Instead, it tracks the combinations that provide the best performance to date. When more combinations are evaluated, the combination becomes the new best combination if the performance of a combination is better than the previous one [22]. This randomness can be more efficient in exploring the hyperparameter space, primarily when enormous search space exists.

### G. Evaluation Model

To evaluate the classification performance of the SSL model, we employ a confusion matrix, as shown in Table II. The confusion matrix will compare the predicted results with the actual class using the rules in Table II. This study uses two parameters for model validation, namely accuracy and F1-score.

TABLE II. CONFUSION MATRIX

		Actual	
		Positive	Negative
Predicted	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative(TN)

Accuracy is the ratio of the correctly predicted dataset to all datasets in the experiment. Accuracy, as in Eq. (4), is a good measurement, but it is only on symmetric datasets, i.e., when the number of false positives and false negatives is almost the same or balanced.

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (4)$$

Precision measures the proportion of genuinely positive instances among all instances predicted as positive by the model. In other words, it quantifies how well the model avoids false positives. High precision indicates that when the model predicts a positive class, it is more likely to be correct. Precision is crucial where the instances of false positives are high or when we want to ensure that the positive predictions made by the model are accurate. However, it is worth noting that precision does not consider instances that were predicted as negative, which could be actual positive cases that were missed (false negatives). Eq. (5) is the precision formula.

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

Recall (sensitivity) is the opposite of precision, as in Eq. (6).

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

Eq. (7) is the average weight of precision and recall. F1 Score is more beneficial than accuracy, especially if the results have an unequal class distribution.

$$F1\ Score = \frac{2(Recall * Precision)}{Recall + Precision} \quad (7)$$

## III. RESULT AND DISCUSSION

### A. Data Distribution

The following Table III shows the distribution of class data for each dataset. Ridife corpus has class positive (24%), neutral (49.1%), and negative (26.9%). IndoNLU Sentiment has class positive (57.7%), neutral (10.7%), and negative (31.6%). Hate Speech only has two classes: positive (42.2%) and negative (57.8%). There are all three datasets in imbalanced class data.

TABLE III. DISTRIBUTION LABEL CLASS

Dataset	Label Class						Total
	Positive	Proportion	Neutral	Proportion	Negative	Proportion	
Ridife	2574	24.0%	5271	49.1%	2882	26.9%	10727
IndoNLU Sentiment	7359	57.7%	1367	10.7%	4034	31.6%	12760
Hate Speech	5561	42.2%	-	-	7606	57.8%	13167

The proposed SSL model employed a tuning parameter. Our experiment uses two techniques, namely Random-Search and Grid-Search. Random Forest and SVM parameters were tuned to gain the best performance of the SSL model. Before applying the SSL model, all datasets corpus are dispart into training, testing, and unlabeled data. Training and testing data are each 10% of the dataset—human labels 10% of training and testing data. The remaining 90% is unlabeled data.

### B. Performance SSL Model

Data testing was employed to assess the performance of the SSL model under both baseline and final conditions. Baseline conditions involved the SSL model being constructed solely using labeled training data. In contrast, the final condition entailed forming the SSL model by fusing labeled training data and Pseudo-Labels generated from unlabeled training data. The experiment was done in two distinct stages. The initial stage entailed evaluating the SSL model using hyperparameters Grid Search, while the subsequent stage involved testing the SSL model performance with hyperparameters using Random Search.

1) *Performance SSL using SVM*: The performance of SSL without hyperparameter is shown in Table IV, while after tuning is shown in Tables V and VI.

TABLE IV. PERFORMANCE SSL SVM

Datasets	Without Hyperparameter			
	Accuracy (%)	Precision (%)	Recall (%)	F1-Score(%)
Ridife	26.9	100	26.9	42.3
IndoNLU Sentiment	57.6	100	57.6	73.1
Hate speech	57.8	100	57.8	73.2

TABLE V. PERFORMANCE SSL SVM TUNED BY GRID SEARCH

Datasets	Grid Search			
	Accuracy (%)	Precision (%)	Recall (%)	F1-Score(%)
Ridife	50.0	97.8	50.0	65.2
IndoNLU Sentiment	78.4	85.5	78.4	81.4
Hate speech	73.7	82.6	73.7	75.6

TABLE VI. PERFORMANCE SSL SVM TUNED BY RANDOM SEARCH

Datasets	Random Search			
	Accuracy (%)	Precision (%)	Recall (%)	F1-Score(%)
Ridife	50.0	98.0	50.0	65.3
IndoNLU Sentiment	78.8	85.2	78.8	81.3
Hate Speech	73.0	83.5	73.0	75.4

Table IV shows the result of the SSL model using SVM with standard parameters having low performance. Accuracy towards all three datasets is under 60%, even though the F1-Score is high. The F1 score shows that the SSL SVM model without parameter tuning is quite good, although only for some data (IndoNLU sentiment and Hate Speech). The Ridife dataset does not reach 50% accuracy, requiring further investigation of the condition of the data, especially the validity of the sample annotation results by Indonesian language experts. With perfect precision values, it is surprising that the recall values are so much lower. With these results, an analysis can be drawn that the standard parameters used in SVM cannot optimize the performance of the SSL model.

Meanwhile, Table V shows that SVM tuning using Grid Search can increase accuracy, indicated by increased accuracy and F1 scores. Because of the differences in class distribution, the performance observations emphasize the F1 score. By comparing the performance of the models without tuning and with tuning, Grid Search increases the F1 Score in the three datasets by 22.9% in the Ridife dataset, 8.3% in the IndoNLU sentiment dataset, and 2.4% in the Hate Speech dataset. Random Search increased the F1 Score by 23.0%, 8.2%, and 2.2%, respectively, on Ridife, IndoNLU sentiment, and Hate speech.

The effect of the parameter tuning done for SVM can be seen in the F1 score performance, as shown in Fig. 3. The F1 score is a valid evaluation to represent the unequal class distribution. Fig. 3 shows that both tuning techniques (Grid and Random Search) can improve the SSL model performance.

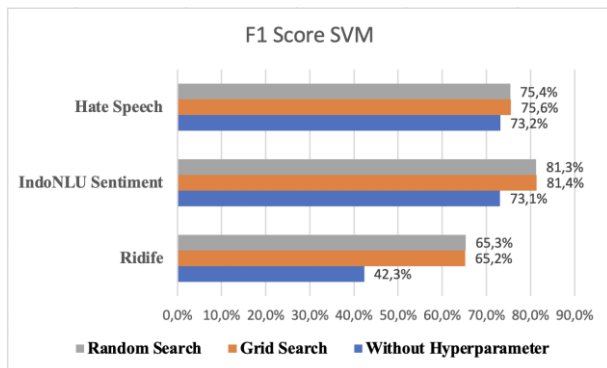


Fig. 3. Performance F1 Score SVM.

2) Performance SSL using random forest (RF): The performance of SSL RF without a hyperparameter is shown in Table VII; SSL tuned by Grid is in Table VIII, and tuned by Random Search is in Table IX. In contrast, tuning on RF does not improve the SSL model's performance like tuning on SVM.

TABLE VII. PERFORMANCE SSL RF

Datasets	Without Hyperparameter			
	Accuracy (%)	Precision (%)	Recall (%)	F1-Score(%)
Ridife	31.5	75.2	31.5	37.1
IndoNLU Sentiment	75.0	81.0	75.0	76.4
Hate Speech	72.8	85.0	72.8	75.5

TABLE VIII. PERFORMANCE SSL RF TUNED BY GRID SEARCH

Datasets	Grid Search			
	Accuracy (%)	Precision (%)	Recall (%)	F1-Score(%)
Ridife	46.0	60.0	46.0	51.0
IndoNLU Sentiment	75.4	80.9	75.4	76.7
Hate Speech	72.6	86.2	72.6	75.6

TABLE IX. PERFORMANCE SSL RF TUNED BY RANDOM SEARCH

Datasets	Random Search			
	Accuracy (%)	Precision (%)	Recall (%)	F1-Score(%)
Ridife	36.5	71.2	36.5	40.7
IndoNLU Sentiment	75.6	81.1	75.6	76.9
Hate Speech	72.0	85.7	72.0	75.0

Table VII shows that the SSL model using RF with standard parameters performs poorly for the Ridife dataset, which only reached 40.7%. However, the IndoNLU Sentiment and Hate Speech datasets are pretty good, above 75%. Meanwhile, Table VII shows that tuning parameter RF using Grid Search was unsuccessful enough to increase model performance. It can be seen that F1 Score before and after tuning in the IndoNLU Sentiment and Hate Speech datasets. By comparing the performance of the RF model without tuning and tuning with Grid Search, the improving performance in all datasets is 13.9% for the Ridife dataset: 0.3% for the IndoNLU Sentiment dataset, and 0.1% for the Hate Speech dataset. While tuning with Random Search could not improve RF performance significantly. The effect of Random Search tuning on RF is only shown by the Ridife dataset. Random Forest is a tree-based algorithm that uses an ensemble tree to increase performance. Therefore, setting up RF with Random Search did not work significantly. The graphical visualization in Fig. 4 supports this. Suppose the conditions of the initial data are observed in more detail. In that case, the Ridife dataset tends to have a lot of noise, slang words, and inaccurate labeling by experts, which may be good for examining the effect of tuning. Meanwhile, the other two datasets (IndoNLU Sentiment and Hate Speech) are relatively cleaner. Considering the condition of the dataset, tuning is suitable for models built from a lot of noise-training data.

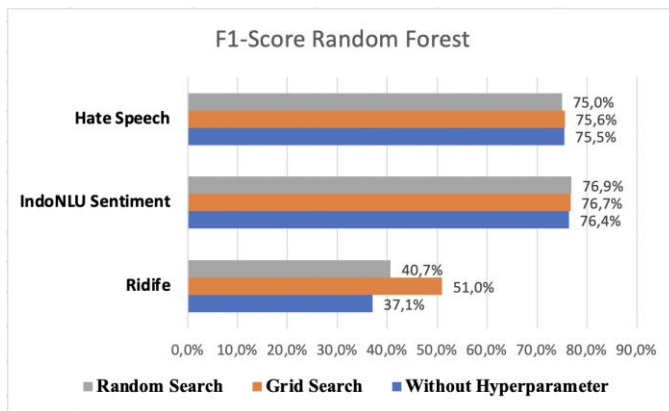


Fig. 4. Performance F1 score RF.

We find that our SSL model for the hate speech dataset obtained higher accuracy (more than 75%) than the SSL proposed by Nur Heri Cahyana et al. [6], which has only reached an accuracy of 59.68% using KNN. Our proposed method proved that SVM and Random Forest perform better for hate speech datasets. In initial data conditions, the Ridife dataset contains more noise, slang words, and inaccurate labels from experts. In contrast, the IndoNLU Sentiment and Hate speech datasets have less noise, and expert labeling is precise. This research found that the proposed SSL using hyperparameter tuning is more suitable for noisy datasets. Hyperparameter tuning is also robust to training data that contains many manual labeling errors by experts.

#### IV. CONCLUSION

Annotation or data labeling in sentiment analysis is a substantial stage in the case of numerous large datasets. Annotation is time-consuming if humans do it. Thus, building model annotation using a computer is needed, but the accuracy model is notable. This research uses a semi-supervised model for annotating sentiment using a Support Vector Machine (SVM) and a Random Forest (RF) algorithm. SVM and RF were respectively tested as classifiers. To gain the most accurate model, RF and SVM were tuned using Random-Search and Grid-Search, respectively. The experiment used three Indonesian corpora as a dataset (Ridife, IndoNLU Sentiment, and Hate speech). Overall, Grid-Search and Random Search leverage performance only in the Ridife dataset. The result shows that tuning works significantly on SVM, but on RF, it does not work on all datasets. This research found that models with hyperparameter tuning are robust to training data containing a lot of noise and incorrect human labeling. For further experiments, employing many variation datasets and paying attention to imbalanced and noise conditions in the data are suggested.

#### ACKNOWLEDGMENT

The authors are grateful to The Ministry of Education, Culture, Research and Technology, Indonesia, for funding this research through the Fundamental Research Grant 2023, which has led to the publication of this paper.

#### REFERENCES

- [1] V. L. S. Lee, K. H. Gan, T. P. Tan, and R. Abdullah, "Semi-Supervised Learning for Sentiment Classification using Small Number of Labeled Data," in *The Fifth Information Systems International Conference*, Surabaya: Elsevier B.V., 2019, pp. 577–584. doi: 10.1016/j.procs.2019.11.159.
- [2] R. Alahmary and H. Al-Dossari, "A semiautomatic annotation approach for sentiment analysis," *J Inf Sci*, 2021, doi: 10.1177/01655515211006594.
- [3] N. H. Cahyana, S. Saifullah, Y. Fauziah, A. S. Aribowo, and R. Drezewski, "Semi-Supervised Text Annotation for Hate Speech Detection using K-Nearest Neighbors and Term Frequency-Inverse Document Frequency," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 10, pp. 147–151, 2022.
- [4] A. Al-Laith, M. Shahbaz, H. F. Alaskar, and A. Rehmat, "Arasencorpus: A semi-supervised approach for sentiment annotation of a large arabic text corpus," *Applied Sciences (Switzerland)*, vol. 11, no. 5, Mar. 2021, doi: 10.3390/app11052434.
- [5] C. R. Aydin and T. Gungör, "Sentiment Analysis in Turkish: Supervised, Semi-Supervised, and Unsupervised Techniques," 2021. doi: 10.1017/S1351324920000200.
- [6] K. Miok, G. Pirs, and M. Robnik-Sikonja, "Bayesian Methods for Semi-supervised Text Annotation," 2020. [Online]. Available: <http://arxiv.org/abs/2010.14872>
- [7] N. H. Cahyana, S. Saifullah, Y. Fauziah, A. S. Aribowo, and R. Drezewski, "Semi-Supervised Text Annotation for Hate Speech Detection using K-Nearest Neighbors and Term Frequency-Inverse Document Frequency," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 10, pp. 147–151, 2022.
- [8] H. J. P. Weerts, A. C. Mueller, and J. Vanschoren, "Importance of Tuning Hyperparameters of Machine Learning Algorithms," Jul. 2020, [Online]. Available: <http://arxiv.org/abs/2007.07588>
- [9] R. Ghawi and J. Pfeffer, "Efficient Hyperparameter Tuning with Grid Search for Text Categorization using kNN Approach with BM25 Similarity," *Open Computer Science*, vol. 9, no. 1, pp. 160–180, 2019, doi: 10.1515/comp-2019-0011.
- [10] L. Villalobos-Arias, C. Quesada-López, J. Guevara-Coto, A. Martínez, and M. Jenkins, "Evaluating hyper-parameter tuning using random search in support vector machines for software effort estimation," in *PROMISE 2020 - Proceedings of the 16th ACM International Conference on Predictive Models and Data Analytics in Software Engineering, Co-located with ESEC/FSE 2020*, Association for Computing Machinery, Inc, Nov. 2020, pp. 31–40. doi: 10.1145/3416508.3417121.
- [11] R. Turner et al., "Bayesian Optimization is Superior to Random Search for Machine Learning Hyperparameter Tuning: Analysis of the Black-Box Optimization Challenge 2020," in *Proceedings of Machine Learning Research*, 2021, pp. 3–26.
- [12] A. Nugroho and H. Suhartanto, "Hyper-Parameter Tuning based on Random Search for DenseNet Optimization," in *7th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*, IEEE Xplore, 2020, pp. 96–99. doi: 10.1109/ICITACEE50144.2020.9239164.
- [13] L. Yang and A. Shami, "On hyperparameter optimization of machine learning algorithms: Theory and practice," *Neurocomputing*, vol. 415, pp. 295–316, 2020, doi: 10.1016/j.neucom.2020.07.061.
- [14] E. S. Tellez, D. Moctezuma, S. Miranda-Jiménez, and M. Graff, "An automated text categorization framework based on hyperparameter optimization," *Knowl Based Syst*, vol. 149, pp. 110–123, 2018, doi: 10.1016/j.knsys.2018.03.003.
- [15] Y. Xie, C. Zhu, W. Zhou, Z. Li, X. Liu, and M. Tu, "Evaluation of machine learning methods for formation lithology identification: A comparison of tuning processes and model performances," *J Pet Sci Eng*, vol. 160, pp. 182–193, Jan. 2018, doi: 10.1016/j.petrol.2017.10.028.
- [16] E. Elgeldawi, A. Sayed, A. R. Galal, and A. M. Zaki, "Hyperparameter Tuning for Machine Learning Algorithms Used for Arabic Sentiment Analysis," *Informatics*, vol. 8, no. 79, pp. 1–21, 2021.

- [17] Md Riyad Hossain and Douglas Timmer, "Machine learning model optimization with hyper-parameter tuning approach," in *International Conference on Advanced Engineering, Technology and Applications (ICAETA)*, 2021. [Online]. Available: <https://www.researchgate.net/publication/354495368>
- [18] S. George and B. Sumathi, "Grid Search Tuning of Hyperparameters in Random Forest Classifier for Customer Feedback Sentiment Prediction," 2020. [Online]. Available: [www.ijacsa.thesai.org](http://www.ijacsa.thesai.org)
- [19] B. H. Shekar and G. Dagnev, "Grid search-based hyperparameter tuning and classification of microarray cancer data," in *2019 2nd International Conference on Advanced Computational and Communication Paradigms (CACCP)*, IEEE, 2019, pp. 1–8. doi: 10.1109/ICACCP.2019.8882943.
- [20] R. Turner *et al.*, "Bayesian Optimization is Superior to Random Search for Machine Learning Hyperparameter Tuning: Analysis of the Black-Box Optimization Challenge 2020," in *Proceedings of Machine Learning Research*, 2020, pp. 3–26.
- [21] R. G. Mantovani, A. L. D. Rossi, J. Vanschoren, B. Bischl, and A. C. P. L. F. De Carvalho, "Effectiveness of Random Search in SVM hyperparameter tuning," *Proceedings of the International Joint Conference on Neural Networks*, vol. 2015-Septe, 2015, doi: 10.1109/IJCNN.2015.7280664.
- [22] S. Andradóttir, "A Review of Random Search Methods," in *Handbook of Simulation Optimization*, M. C. Fu, Ed., New York, NY: Springer New York, 2015, pp. 277–292. doi: 10.1007/978-1-4939-1384-8\_10.

# Usability Testing of Memorable Word in Security Enhancing in e-Government and e-Financial Systems

Hanan Alotaibi, Dania Aljeaid, Amal Alharbi  
Faculty of Computing and Information Technology  
King Abdulaziz University  
Jeddah, Saudi Arabia

**Abstract**—Most applications increase their security by adding an extra layer to the login process using two-factor authentication (2FA). In Saudi Arabia, One-Time Password (OTP), which is 2FA, is the most common method used as users log in to their accounts. However, some issues have emerged with using OTP as 2FA; these issues from previous research were investigated in the study. Also, the study proposed a new method of account authentication, which is a Memorable Word (MW). MW is the second and short password in which the user enters a certain number of characters instead of the whole password. The study conducted usability testing to compare two 2FA methods, OTP and MW. The study included 60 participants logged into a simulated website using both authentication methods. Then, all participants have to complete the questionnaire. The collected data analyses showed a favourable opinion of the MW method.

**Keywords**—Security; usability testing; two factor authentication; one time password; memorable word

## I. INTRODUCTION

The digital world is a rapidly evolving landscape, and using e-Systems is becoming increasingly commonplace. e-Systems are electronic systems that allow users to access, store, and share information and data. These systems are used in various ways, from e-government and online banking to social media, and they are becoming an integral part of our lives. e-Government has the potential to modernise the way governments interact with citizens and provide services. It can give citizens access to government services and information more efficiently and cost-effectively [1]. It can also help governments to manage their resources better and improve the delivery of services. e-Government can help reduce the administrative burden on government employees and improve the transparency and accountability of government services [2, 3, 4]. For instance, the government of Saudi Arabia has taken a proactive approach to ensure the security of its e-government systems. The Saudi e-government security policy is designed to protect data resources from a wide range of risks, including malicious attacks, unauthorised access, and data loss, by implementing aspects of the data security [5].

Similarly, e-financial services in Saudi Arabia have seen a significant shift in recent years as banks have offered more innovative services through online banking. This shift has enabled banks to maintain their market share and gain customers as online banking has become increasingly popular [6]. However, this digital transformation to systems poses several challenges and security threats. For example, it can be

difficult to ensure that systems are reliable and resilient in the face of cyber-attacks and other threats.

However, with the increased use of systems comes an increased risk of security threats. Cybercriminals are constantly looking for ways to exploit weaknesses in these systems and expose sensitive information such as social security numbers or verification numbers sent to e-mail addresses or contact numbers [7, 8]. Cybercriminals use a variety of methods to accomplish successful cyberattacks, which include phishing, malware, and ransomware attacks, as well as data breaches. To protect against these threats, it is essential to implement strong security measures in place within systems. This includes keeping software and systems up to date, which can help prevent vulnerabilities from being exploited. It is essential to be aware of the latest security threats and to take steps to protect against them. Moreover, using strong passwords, multifactor authentication, and encryption are essential fields [9]. An authentication system is deployed to ensure that both parties involved in the communication are authentic ones. The dominant form used in various authentication systems is based on username and password. Nevertheless, passwords can be easily guessed if it is weak or stolen, so it is significant to use additional security measures. One of the most common methods used to increase the security of authentication is to adopt two-factor authentication (2FA). Implementing 2FA for end users can provide organisations with several benefits, including increased security and improved user experience. Recent studies found that 2FA can help protect user accounts from unauthorised access and malicious actors and reduce the risk of data breaches [10, 11]. However, it can also pose some challenges and complications. For example, implementing 2FA can be difficult and costly, as organisations need to invest in the necessary infrastructure and technologies to support the authentication process. Additionally, users may find entering their 2FA code tedious and time-consuming [10, 12]. Thus, this research studies the implementation of Memorable Word (MW) as a 2FA and evaluates its security efficiency and usability.

The main contribution of this paper is to investigate the users' perspective when accessing Saudi systems, such as government and bank websites, using MW instead of the current authentication method, SMS One-Time Password (OTP).

The rest of the paper is organised as follows: Sections II describes the common methods used in authentication while Section III highlights recent works related to usability testing

in authentication, followed by methodology in Section IV. In Section V, exhaustive experiments are conducted to validate the MW. In Section VI, the results of the experiments are discussed. Lastly, the paper is concluded along with the limitations and future work in Sections VI and VIII, respectively.

## II. AUTHENTICATION METHODS

Due to the increasing number of cybercrimes, traditional username and password authentication methods are no longer enough to protect sensitive information from malicious actors. Businesses and individuals must take additional steps to secure their data. Multi-factor authentication (MFA) is one of the most effective methods for protecting sensitive information. MFA requires users to provide two or more pieces of evidence to prove their identity. This could include a combination of something the user knows (such as a password), something the user has (such as a physical token or smartphone), and something the user is (such as a biometric scan) [13]. 2FA is one of the most common authentication methods. It is a security measure that requires two different authentication factors to verify a user's identity. It is used to protect sensitive information from unauthorised access and is becoming increasingly popular as a way to protect against data breaches. The two factors used in 2FA are typically something the user knows (e.g., a password or PIN) and something the user has (e.g., an OTP, token, or digital certificate). This combination of factors makes it much more difficult for an attacker to gain access to a user's account, as they would need to know both the password and have access to the physical device or token. 2FA is a great way to protect against phishing, brute force attacks, keylogging, and credential theft attacks. It is also a great way to protect against data breaches, requiring two different authentication factors to verify a user's identity. There are several different types of 2FA, such as OTP-based and Biometrics-based. OTP-based 2FA requires users to enter an OTP valid for a single login session. Biometrics-based 2FA requires a user to provide a biometric factor such as a fingerprint or iris scan [6]. Several different 2FA methods exist, such as SMS, Time-Based One-Time Password, Pre-generated Codes, Push, and Universal Second Factor Security Keys. Each method has advantages and disadvantages, as discussed in the following subsections.

### A. SMS Token / SMS-based Authentication

SMS-based authentication is one of the most common methods of 2FA, and many organisations use it to protect their users' accounts and systems. In SMS-based, a one-time verification code is sent to the user via a text message to their mobile phone. This code is usually six digits long and is used to verify the user's identity. The user then enters the code into the system to gain access. This code is only valid for a short period of time, usually a few minutes, and it must be used within that time frame, or it will expire. It is easy to use and requires minimal effort from the user [6, 14]. However, SMS-based 2FA is not without its drawbacks. It is vulnerable to SIM-swapping attacks, where an attacker can access the user's phone number and intercept the verification code. It is also vulnerable to phishing attacks, where an attacker can send a fake text message with a malicious link that leads to a fake

website. SMS messages can take a long time to arrive, and they can be blocked or delayed by network congestion because they are delivered over cellular network standard SMS. This can be a problem if users need to access their accounts quickly [15, 14]

### B. Time-Based One-Time Password

Time-based one-time password (TOTP) is an alternative to SMS-based 2FA that provides an additional layer of security for online accounts. It generates a unique, valid code for a limited time, usually 30 seconds. This code then authenticates the user's identity [6, 16]. Yet, there are some drawbacks to using it. One of the main drawbacks is that it can be challenging to use. The user must have access to the device generating the code, such as a smartphone or a hardware token. This can be inconvenient for users who do not have access to the device or who do not have the time to wait for the code to be generated. Another drawback is that TOTP codes can be vulnerable to replay attacks. This is when an attacker captures the code and uses it to gain access to the account. To mitigate these attacks, it is essential to use a secure connection when generating the code and ensure that it is not stored in plain text. Finally, TOTP codes can be challenging to remember. This can be a problem for users not using 2FA [17].

### C. Pre-Generated Codes

Pre-generated tokens are an effective backup 2FA method if the user cannot access the primary 2FA method. This method is relatively straightforward to implement, as the service provider simply creates a list of verification codes and asks the user to print or write down the codes. The list length is variable; the codes are usually about eight digits long. Tokens can be used in any order and must be kept secure by both the server and the user to prevent theft. Since these codes are usually longer than codes sent via SMS or generated using TOTP, there is additional room for user error when entering codes. Moreover, the user must be careful not to lose the broker on which they registered the codes, and they will be vulnerable to an offline brute force attack [6].

### D. Push

Push authentication requires the user to receive a push notification on their smartphone to approve or deny a login attempt. This method is advantageous because it eliminates the need for users to type in numbers, as required by other 2FA methods, making it both faster and more user-friendly [18]. Additionally, push authentication requires Internet access, which is necessary to keep communication between the user's device and the server secure, such as through TLS. However, the most prominent push-based authentication methods are proprietary, making it difficult to verify the exact security measures in place and require implicitly trusting a third party [6]. Furthermore, push-based authentication has not yet been well-studied by the security community, making it difficult to assess the security of this method.

### E. Universal Second Factor Security Keys

Universal Second Factor (U2F) Security Keys are an open standard for authentication through a USB device. The user must connect the device to the computer to authenticate with a security key and activate the device when the website requests.



U2F Security Keys are designed to be more secure than traditional 2FA methods, such as SMS or email-based authentication. One of the main drawbacks of U2F security keys is that they can be challenging to set up and use. U2F security keys require users to install a particular driver on their computer or mobile device to use them. This can be a time-consuming process, and it can be difficult for users who need to be tech-savvy. Additionally, U2F security keys are incompatible with all devices and can be expensive. Besides, U2F security keys can be lost or stolen, which can be a significant inconvenience for users who need to replace them [19].

#### F. Memorable Word Technique

The Memorable Word (MW) is a short password (usually assumed to consist of one word) and is one of the layers of authentication, where the MW differs from the password in how it is used; instead of entering the whole MW, the user enters a certain number of MW's characters, usually, three letters and the letters to be entered vary each time the user is asked to enter it. Initially, the client and the server know a short password of length  $m$  characters that has been shared before. During authentication, the server sends unique numbers between 1 and  $m$  to the client. The client responds with the letter in each corresponding position in the password. If all these characters are correct, the authentication succeeds; otherwise, it fails [20]. It has been recognised that transaction systems must include authentication and data encryption. To complete these requirements, MW has been proposed to contain mutual authentication and data encryption using the symmetric algorithm that improves the security of existing transactions. Symmetric encryption is employed in transactions to prevent identity theft of clients or banks—additionally, user authentication and authorisation to protect against cyber-attacks [21, 22].

### III. RELATED WORK

Usability testing can help identify areas for improvement in the system, such as user experience and security, which can help improve the system's overall performance. Furthermore, usability testing can help to identify any areas of confusion or difficulty that users may need help with when using the 2FA system. This can help improve the user experience by making the system easier and more convenient. Usability testing is one of the preferred methods for assessing user experience with 2FA due to its ability to provide direct feedback from users [11]. Das et al. [23] suggest that usability testing can be used to evaluate the effectiveness of 2FA systems and identify potential usability issues. This type of testing often includes a series of tasks to measure user performance and satisfaction with the 2FA process. Other evaluation methods, such as surveys, interviews, and focus groups, provide additional insight into user experience with 2FA [11]. Gunson et al. [24] found that usability testing was the most effective way to evaluate 2FA due to the complexity of the task and the need to ensure that users understand and follow the authentication process correctly. However, they also identified several challenges in implementing usability testing. These include finding a representative sample of users, developing suitable test scenarios, and ensuring that the test results are valid and

reliable. According to a study by Golla et al. [24], implementing 2FA for end users can provide organisations with several benefits, including increased security and improved user experience. The study found that 2FA can help protect user accounts from unauthorised access and malicious actors and reduce the risk of data breaches. Additionally, 2FA can provide users with an improved experience when logging into their accounts, as they can be quickly and easily authenticated. However, while 2FA can provide organisations and end users many benefits, it can pose some challenges. The study found that implementing 2FA can be difficult and costly, as organisations need to invest in the necessary infrastructure and technologies to support the authentication process. Additionally, users may find entering their 2FA code tedious and time-consuming [24]. As such, organisations must be aware of these challenges and take steps to ensure that their implementations of 2FA are secure, efficient and user-friendly. Abbott and Patil [12] explored the potential for improving user experience through 2FA. They found that 2FA can enhance security, provide users with greater confidence in the service, and provide a better overall user experience. Through their study, the authors found that users are more likely to remain engaged with services that use 2FA as they feel more secure and trust the service provider. Furthermore, they noted that 2FA can be tailored to the individual user's needs, which can help to improve user experience by providing them with a more tailored experience.

The usability dimensions of ISO 9241-11 form the basis for measuring user experience based on the three usability dimensions, which are efficiency, effectiveness, and satisfaction. Efficiency focuses on the amount of time taken to complete a task, effectiveness is the ability to complete a task, and satisfaction is the user's opinion of the convenience and acceptability of the system, which is measured by the System Usability Scale (SUS) [25]. Factors impacting each usability dimension can be documented, such as the time it takes to complete a task or user demographic information. Collecting this information can help developers and designers create and improve the 2FA MW systems [26]. SUS is a tool that has become increasingly popular in the field of usability testing. Developed by John Brooke in 1986, SUS is an efficient and cost-effective way to measure the usability of a system. The tool has been used in a wide range of studies and has consistently produced reliable results. Additionally, SUS is a popular option for usability testing because it is efficient and cost-effective. It is a straightforward tool that can be administered quickly and easily.

Weir et al. [27] compared the usability of three two-factor authentications: push-button tokens, card-activated tokens, and PIN-activated tokens. The study aimed to measure the time required for authentication and user satisfaction. Their findings were that user prefers authentication methods which are easy to use rather than security; however, quality and usability decreased when additional levels of security were required.

An exploratory comparative investigation study conducted by [28] into the usability of 2FA. The study assessed and compared the usability of three widely used 2F solutions: security token-generated codes, OTP delivered through email or SMS, and dedicated smartphone apps like Google

Authenticator. Also, they investigated motivations behind users' choices and examined how these factors influence their perception of usability. The finding from the study indicates that 2FA are widely accepted and highly usable. This means that users are embracing the user of additional layer of security beyond just password. The study also highlights that user opinions of 2F usability are frequently connected with individual attributes. This suggests that different users may have varying levels of comfort or experience with using 2FA, which can influence their perception of its usability. Moreover, the study reveals that the trustworthiness of 2F is positively correlated with ease of use.

In a study conducted by Reese et al. [29] the usability of five 2FA methods was examined. The study aimed to compare the usability of Pre-generated Codes, Push, SMS, OTP, and U2F Security Keys. The sample consisted of 72 participants who logged into a simulated banking website using 2FA. The objective was to gain insights into users' perspectives on 2FA and assess their experiences with different authentication methods. The findings showed that the majority of participants expressed a desire to use 2FA as a means to enhance the security of their sensitive online accounts. However, it was noted that some participants encountered difficulties when utilizing these methods such as spending longer time at login phase, particularly with the OTP and U2F methods.

Our research is different from the previous studies as it focuses on the implementation of a new two-factor authentication (2FA) method, specifically the Memorable Word (MW), in Saudi Arabia's e-government and banking systems. The aim of our study was to examine the user perspective on utilizing MW as a 2FA method. To achieve this objective, a survey was conducted to compare the currently employed 2FA method in Saudi Arabia, which is One-Time Password (OTP), with the proposed 2FA MW method.

#### IV. METHODOLOGY

The study's first phase, as described in Sections II and III, involved conducting a literature review on authentication methods and usability. By thoroughly examining existing literature, gaps in knowledge were identified, highlighting areas that required further research. In the second phase, a website was developed to support both OTP-2FA and MW methods. To gain a deeper understanding of MW authentication, a questionnaire was conducted through a simulated webpage. Additionally, three expert reviews were obtained to ensure that the usability testing aligned with the study's objectives. The purpose of these interviews was to assist the authors in confirming the usability testing process. Afterward, a comprehensive questionnaire was constructed and administered to collect data on users' perspectives towards both OTP and MW methods. The research methodology, including the various phases and steps undertaken, is presented in Fig. 1.

The participants were assigned various tasks to perform on a simulated website and subsequently completed a survey. To ensure a smooth process, participants were initially scheduled for an appointment with a study coordinator. During this meeting, the coordinator provided necessary guidance and assistance to the participants in creating an account as the following steps:

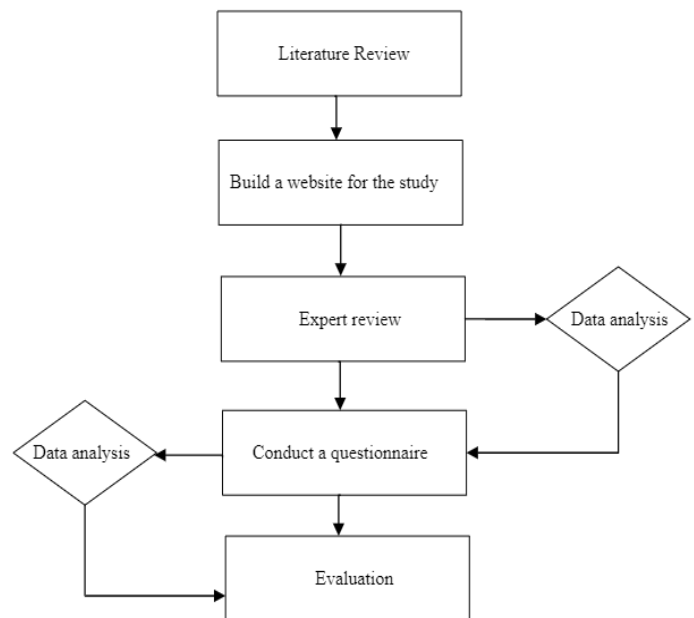


Fig. 1. Research methodology.

1) *User registration*: This step involved user registration, where the participants were required to provide their details, such as their name, phone number, and email address. Once registered, the participants were able to create their unique username and password. In addition, they were required to set up an MW with the following considerations:

- Six to eight characters long without spaces.
- Contains only letters (A to Z), excluding any numbers or special characters.
- It cannot be the user's first and last name.
- It should not include alphabetic sequences such as "QWERT" keyboard.

The participants were informed that they would need to remember three random characters from their MW, which they would be prompted to enter after providing their login credentials. Fig. 2 shows the creation MW page.

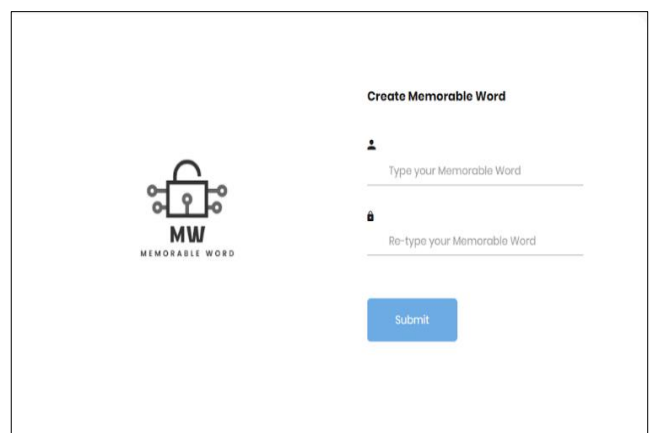


Fig. 2. Create MW page.

2) *User log-in*: The participants were prompted to log in to the website after successfully registering. To ensure that the participants were familiar with both authentication methods, they were instructed to initially select the OTP method and then log out and re-login using the MW method. This step was deemed necessary as it allowed the users to experience and become comfortable with both methods before proceeding to the next stage. Fig. 3 shows the login page, where the user can select the authentication methods. Fig. 4 shows the MW authentication page.

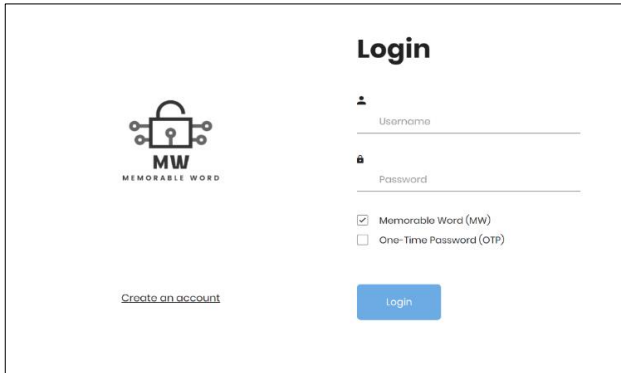


Fig. 3. Login page.

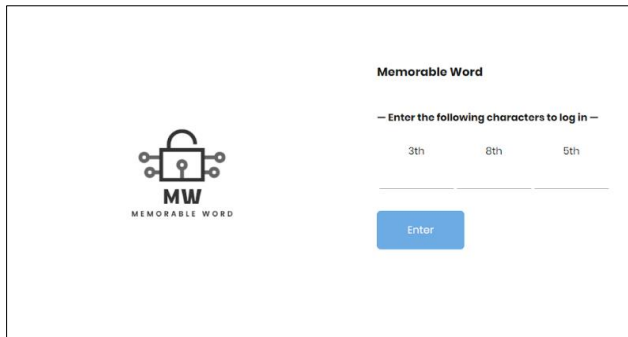


Fig. 4. The MW authentication page.

3) *Survey questionnaire*: Once the participants had completed the registration and login process, they were directed to answer a questionnaire. A total of 60 participants were recruited for this purpose, and they were randomly selected to ensure unbiased results. After two weeks, the collected data from the questionnaire was meticulously analysed, providing valuable insights and conclusions for further evaluation.

## V. RESULTS ANALYSIS

The data was collected from 60 participants in a span of two-week. In this study, participants were asked to answer eight parts of a survey, each consisting of questions with different objectives.

### A. Demographics/ Participants

The demographic data collected from the participants revealed a slightly higher number of female participants than male participants, with 62.5% and 37.5%, respectively.

Additionally, the study showed that most participants were young adults, with 65.5% being between the ages of 18-29 years, 23.6% between 30-49 years, and only 10.9% between 50-69 years. The data also indicated that most participants had a bachelor's degree, accounting for around two-thirds of the participants (63.6%). Interestingly, all participants were familiar with using e-government and online bank systems, indicating their technological proficiency. However, it was also found that 100% of the participants had no prior knowledge of the MW method. To ensure that participants clearly understood the MW method, a brief description was given to them during the registration process through the simulated website.

Table I summarises the participant's demographics. As can be seen from the table, although the sample size is small (total = 60), overall, the study showed that the participants were diverse in gender, age, and education level, but their technological proficiency was high.

TABLE I. PARTICIPANTS' DEMOGRAPHICS (TOTAL = 60)

Gender	
Male	37.5%
Female	62.5%
Age	
18-29 years	65.5%
30-49 years	23.6%
50-69 years	10.9%
Qualification	
Secondary School	20%
Diploma	7.3%
Bachelor	63.6%
Postgraduate	9.1%
Familiar with using E-government and online bank systems?	
Yes	100%
No	0%
Familiar with Memorable Word?	
Yes	0%
No	100%

### B. Timing Data

Login timing is an essential element in analysing user experience regarding systems. The study analysed the timing data of two authentication methods used for logging into a system: OTP and MW. The login time for two different methods was measured, counting the time from when the login page initially loaded to when the user submitted a password. To obtain reliable data, users were asked to repeat the login process ten times, five times for each method. Once all the data was collected and analysed, the findings revealed that users spent more time in the OTP method due to the delay in receiving the verification code. Fig. 5 presents the time spent in second to login to the system using both methods. Table II shows the mean time in seconds for both methods.

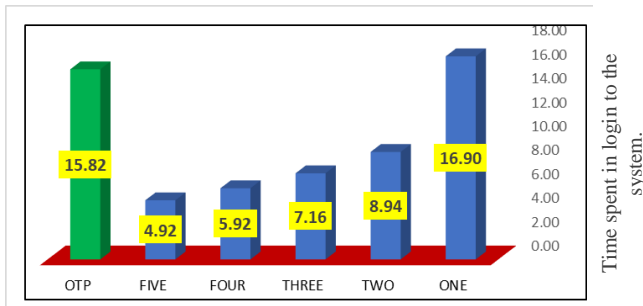


Fig. 5. The time spent to login to the system using both methods.

TABLE II. THE MEAN TIME IN SECONDS FOR THE TWO AUTHENTICATION METHODS

Method	Mean Time
MW	9.3
OTP	15.81

### C. Individual Learnability

One of the objectives of the survey was to explore individual learnability. The hypothesis in this part of the experiment was that participants would become faster at validating their accounts with specific authentication methods as they become more familiar with them. To test this hypothesis, we computed a correlation between the time an individual spent in the session and the amount of time it took to validate their account using different authentication methods. The results showed a statistically significant difference between the two authentication methods (P-value  $\leq 0.01$ ). The MW method was found to be faster than the OTP method. This suggests that individuals can learn and become faster at validating their accounts with MW authentication methods over time.

### D. System Usability Scale

The main goal of SUS is to evaluate a user's perception of a system's ease of use and overall usability. The SUS survey consists of nine tool questions with five-point Likert-scale answers. The survey is designed to gather feedback from users about how easy it is to use an MW authentication. To increase the reliability of the survey, four of the questions are phrased positively, while the other five are phrased negatively. This approach helps to reduce response bias and provides more accurate results. The results indicated that 98.18% of participants found the MW method easy to use for login, demonstrating high usability. Additionally, 94.91% of participants preferred the MW method to the OTP method. These results were all statistically significant and supported the hypothesis that the MW method was easy to use and preferred by users (P-value = 0.00  $\leq 0.05$ ). Table III and Fig. 6 show the findings from each question.

### E. Previous Experiences with Account Compromise and Worth Inconvenience

Participants were asked if they had ever faced difficulty logging in to their accounts regarding compromised online accounts. 12.7% of participants have an experience with remote attackers taking over their online accounts. Also, 52.7% know someone had an experience with remote attackers taking

over their online accounts. Participants with previous experience and who know someone with an account compromised would be more likely to feel that an MW method was worth using. For worth inconvenience, the participants were asked to use MW as a second authentication is worth an extra step of login. 87.27% of participants felt using MW is worth an additional inconvenience.

### F. Security and Inconvenience

The security and inconvenience factor was also investigated. Participants were asked if the MW authentication method made them feel more efficient and convenient to access their accounts than the OTP method. Most participants thought that the MW method was secure and convenient. The study showed that 92.36% of participants felt that MW was secure when logging into their accounts. Furthermore, 89.45% of participants found MW more convenient than OTP.

TABLE III. SYSTEM USABILITY SCALE ANALYSIS OF THE NINE STATEMENTS

Statement Number	Statement	Mean	Standard Deviation	P-value
1	I find the various functions in this MW were well integrated.	4.95	0.23	.000
2	I think the MW was easy to use.	4.91	0.55	.000
3	I would like to use this MW frequently.	4.75	0.48	.000
4	I think most users would learn to use this MW very quickly.	4.69	0.96	.000
5	I find the MW is complex.	1.51	0.84	.000
6	I find the MW is hard to use.	1.51	0.84	.000
7	I think I need support to use MW.	1.45	0.54	.000
8	I need time to learn how to use MW	1.38	0.97	.000
9	I think there is inconsistency in using MW.	1.11	0.57	.000

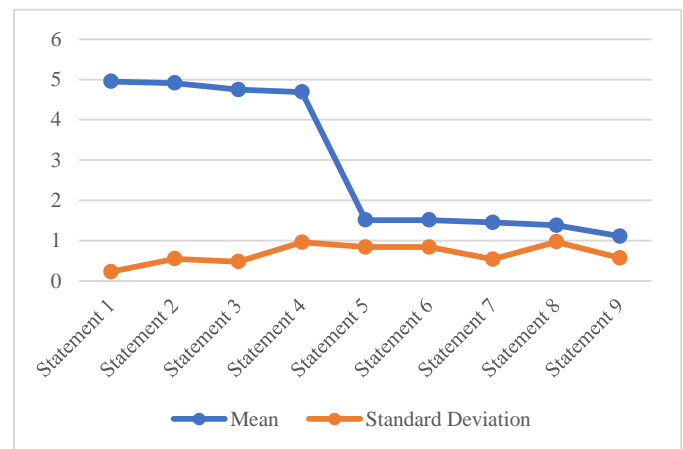


Fig. 6. System usability scale analysis of the nine statements.

### G. Perception of Likelihood for Account Compromise

Participants were asked how much value they placed on their online accounts, such as bank and government accounts, to investigate whether they felt the need to protect their information and data. The results showed that 92.5% of participants expressed that they needed to secure their accounts from others, while 7.5% felt that there was nothing essential to protect.

### H. MW Timeout

After a week of creating an account on the simulated website, the participants were asked to log in to their account using MW as an authentication factor to test their experience using the proposed method if they needed help entering randomly selected letters. The result is that 94.91% of participants logged into their accounts without any mistakes; before time out; also, 97.82% found MW easy to use. 96.36% of participants agreed with the statement, "I did not struggle in using MW as much as struggling in OTP".

## VI. DISCUSSION

The main contribution of this study is to investigate the users' perspective when accessing Saudi electronic systems, such as government and bank websites, using MW instead of the current authentication method, SMS OTP. The study found a favourable opinion and feedback toward MW, where the participants reported faster login times than the OTP method. An overwhelming majority of participants (97.82%) logged into their accounts without any mistakes or forgetting their memorable words, indicating that MW was convenient, learnable, and easy to use.

Comparing the study results with similar study results [29], we found that timing data in MW authentication method is the faster way for users to log in to their accounts. At the same time, U2F is the quickest method in the [29]. Moreover, our study tested MW timeout to check if participants needed help entering MW letters. The result was that 94.91% of participants logged into their accounts without any mistakes before it timed out. Another study tested OTP timeout and found that 65% of participants had problems entering the six-digit verification code before it timed out. Both studies conducted SUS to evaluate a user's perception of a system's ease of use and overall usability. Our study found that the median score of the MW method was 96.22. Regarding Reese et al. study [29], the finding of evaluating five methods was passwords had the highest median SUS score, with a median score of 95, followed by TOTP, which had a median SUS score of 88.75.

In today's fast-paced world, users have increasingly high expectations for carrying out their tasks promptly on various digital platforms. Any delay or inconvenience in the login process can lead to frustration, negative user experiences, and potential cyberattacks. Our study highlighted a critical finding regarding the OTP method, which showed a delay in receiving the verification code. Eliminating the need for users to wait for OTPs can enhance the efficiency of authentication systems and provide a seamless login experience for users. OTP can be less secure when a user's mobile device or token generator is compromised or intercepted maliciously. One of the significant vulnerabilities associated with SMS OTP is SIM swapping [30,

31]. This occurs when an attacker convinces a mobile network provider to transfer a victim's phone number to a new SIM card under their control. Once the attacker controls the victim's phone number, they can intercept any SMS OTPs sent to that number, effectively gaining access to the victim's accounts. Another common attack vector for OTP is the phishing attack [32, 33]. Phishing involves tricking and deceiving the victim into revealing their login credentials or multi-factor authentication (MFA) code by posing as a legitimate entity. In the case of OTP, the attacker could send a phishing message to the victim's phone, making it appear as if it is coming from a trusted source such as a bank or other known service providers. Once the victim falls for the scam, the attacker can use that code to authenticate themselves and gain unauthorised access to the victim's account.

On the other hand, using the MW technique can help prevent certain types of attacks, such as MITM attacks [34, 35], where malicious attackers redirect users to a fake website before forwarding them to the legitimate one. Since users are only required to type random letters from the memorable word instead of their entire secret token, it becomes more difficult for attackers to capture the complete word in one go. While the MW technique may provide some protection against MITM attacks, unfortunately, it does not entirely prevent them. In the event of a MITM attack, an attacker could still prompt the user for the same letters they are being prompted for, thereby gaining access. Thus, it is recommended that the MW be changed frequently.

It is important to consider that while MW may provide some level of security at the user's end, it may be less secure at the server or organisation's end. One potential vulnerability exists in storing memorable words as a single field within the system. This means that if there were to be a database leak or breach, the MW authentication system would be susceptible to the same risks as other forms of 2FA.

In conclusion, using MW as a security solution offers few advantages over OTP methods. It provides a better level of security than just relying on a password alone by preventing the transmission of the complete word and protecting against certain types of attacks. However, it may be less secure in specific scenarios and should ideally be implemented as an additional layer of security alongside other authentication methods.

## VII. LIMITATION

The main limitation of this research was the time needed to spend with the participants, which reduced the sample size. Thus, this may prevent generalising the findings to the general population.

## VIII. CONCLUSION AND FUTURE WORK

The study investigated and compared the usability of two authentication methods, MW and OTP, in e-government and e-bank systems in Saudi Arabia. A usability testing survey was conducted to gain insights into users' perspectives on the proposed MW method and compare it with OTP. Overall, the participants expressed a positive opinion about MW, finding it easy to use and highly convenient. In contrast, OTP presented several challenges, including delays in receiving verification

codes and ineffective authentication when the signal was interrupted. The study revealed that many users struggled with OTP and required an alternative authentication method. Through analysis and simulation, it was determined that the proposed MW method offers comparable security control to existing OTP authorisation while minimising the dynamic risk of theft and eliminating the need for additional hardware. As this method eliminates the possibility of crucial theft, it can be used on private and public computers. Notably, this method is cost-effective and poses no significant hurdles.

The proposed method holds potential for further research on security attacks, particularly in addressing replay attacks, man-in-the-middle attacks, reflection attacks, and parallel session attacks.

#### ACKNOWLEDGMENT

The authors are grateful to Afrah Almalki and Retal Mehdawi for their work during projects.

#### REFERENCES

- [1] A. S. Alharbi, G. Halikias, M. Rajarajan and M. Yamin, "A review of effectiveness of Saudi E-government data security management," *International Journal of Information Technology*, vol. 13, p. 573–579, 2021.
- [2] H. P. Singh and T. S. Alshammari, "An Institutional Theory Perspective on Developing a Cyber Security Legal Framework: A Case of Saudi Arabia," *Beijing Law Review*, vol. 11, no. 3, pp. 637-650, 2020.
- [3] G. P. Dias, "Global e-government development: besides the relative wealth of countries, do policies matter?," *Transforming Government: People, Process and Policy*, vol. 14, no. 3, pp. 381-400, 2020.
- [4] Y.-C. Yan and S.-J. Lyu, "Can e-government reduce local governments' financial deficits?—Analysis based on county-level data from China," *Government Information Quarterly*, vol. 40, no. 3, 2023.
- [5] A. Alrubaq and T. Alharbi, "Developing a Cybersecurity Framework for e-Government Project in the Kingdom of Saudi Arabia," *Journal of Cybersecurity and Privacy*, vol. 1, no. 2, p. 302–318, 2021.
- [6] R. A. Abdulhadi and S. Ahmad, "Internet Banking In Saudi Arabia," *Palarch's Journal Of Archaeology Of Egypt/Egyptology*, vol. 18, no. 13, pp. 673-684, 2021.
- [7] F. Mabrouk, "Statistics of Cybercrime from 2016 to the First Half of 2020," *International Journal of Computer Science and Network*, vol. 9, no. 5, pp. 252-261, 2020.
- [8] M. Bada and J. R. C. Nurse, "Exploring Cybercriminal Activities, Behaviors, and Profiles," in *Applied Cognitive Science and Technology*, Singapore, Springer, 2023, p. 109–120.
- [9] R. Dhanalakshmi, N. Vijayaraghavan, S. Narasimhan and S. Basha, "Password Manager with Multi-Factor Authentication," in *International Conference on Networking and Communications (ICNWC)*, Chennai, 2023.
- [10] M. Golla, G. Ho, M. Lohmus, M. Pulluri and E. M. Redmiles, "Driving 2FA Adoption at Scale: Optimizing Two-Factor Authentication Notification Design Patterns," in *30th USENIX Security Symposium*, 2021.
- [11] K. Reese, "Evaluating the Usability of Two-Factor Authentication," 2018.
- [12] J. Abbott and S. Patil, "How Mandatory Second Factor Affects the Authentication User Experience," in *CHI Conference on Human Factors in Computing Systems*, Honolulu, 2020.
- [13] E. T. Alharbi and D. Alghazzawi, "Two Factor Authentication Framework Using OTP-SMS Based on Blockchain," *Transactions on Machine Learning and Artificial Intelligence*, vol. 7, no. 3, pp. 17-27, 2019.
- [14] R. P. Jover, "Security Analysis of SMS as a Second Factor of Authentication: The challenges of multifactor authentication based on SMS, including cellular security deficiencies, SS7 exploits, and SIM swapping," *acmqueue*, vol. 18, no. 4, p. 37–60, 2020.
- [15] V. K. Anand and D. Tirfe, "A Survey on Trends of Two-Factor Authentication," in *Contemporary Issues in Communication, Cloud and Big Data Analytics*, Singapore, 2022.
- [16] M. Hassan, Z. Shukur and M. K. Has, "An Improved Time-Based One Time Password Authentication Framework for Electronic Payments," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 11, pp. 359-366, 2020.
- [17] A. A. Ali Abdullah S. Alqahtani, "0E2FA: Zero Effort Two-Factor Authentication," 2020.
- [18] A. Mohammed, R. Dziauddin and L. Abdul Latiff, "Current Multi-factor of Authentication: Approaches, Requirements, Attacks and Challenges," *Current Multi-factor of Authentication: Approaches, Requirements, Attacks and Challenges*, vol. 14, no. 1, pp. 166-178, 2023.
- [19] S. Das and A. Dingman, "Why Johnny Doesn't Use Two Factor A Two-Phase Usability Study of the FIDO U2F Security Key," in *Financial Cryptography and Data Security: 22nd International Conference, Nieuwpoort*, 2018.
- [20] D. Tirfe and V. K. Anand, "A Survey on Trends of Two-Factor Authentication," in *Contemporary Issues in Communication, Cloud and Big Data Analytics*, Singapore, 2021.
- [21] S. Istiyaq, "Hybrid Authentication System Using QR Code with OTP," *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, vol. 10, no. 6, pp. 1194-1197, 2016.
- [22] S. P. Boraiah, "Secure Cardless Transaction Android Application using ECC algorithm and QR code," *Masters thesis*, Dublin, National College of Ireland, 2019.
- [23] S. Das, B. Wang, Z. Tingle and L. J. Camp, "Evaluating User Perception of Multi-Factor Authentication: A Systematic Review," in *the Thirteenth International Symposium on Human Aspects of Information Security & Assurance (HAISA 2019)*, Nicosia, 2019.
- [24] N. Gunson, D. Marshall, H. Morton and M. Jack, "User perceptions of security and usability of single-factor and two-factor authentication in automated telephone banking," *Computers and Security*, vol. 30, no. 4, pp. 208-220, 2011.
- [25] J. Brooke, "SUS—A Quick and Dirty Usability Scale," *Usability Evaluation in Industry*, pp. 189-194, 1986.
- [26] A. S. Alharbi, G. Halikias, M. Rajarajan and M. Yamin, "A review of effectiveness of Saudi E-government data security management," *International Journal of Information Technology*, vol. 13, no. 2, pp. 573-579, 2021.
- [27] C. Weir, G. Douglas, M. Carruthers and M. Jack, "User perceptions of security, convenience and usability for ebanking authentication tokens," *Computers & Security*, vol. 28, no. 1-2, pp. 47-62, 2009.
- [28] E. D. Cristofaro, H. Du and J. Freudige, "A Comparative Usability Study of Two-Factor Authentication," in *Workshop on Usable Security and Privacy (USEC'14)*, 2014.
- [29] K. Reese, T. Smith, J. Dutton, J. Armknecht, J. Cameron and K. Seamon, "A usability study of five {two-factor} authentication methods," in *Fifteenth Symposium on Usable Privacy and Security*, Santa Clara, CA, USA, 2019.
- [30] M. Kim, J. Suh and H. Kwon, "A Study of the Emerging Trends in SIM Swapping Crime and Effective Countermeasures," in *2022 IEEE/ACIS 7th International Conference on Big Data, Cloud Computing, and Data Science (BCD)*, Vietnam, 2022.
- [31] R. P. Jover, "Security analysis of SMS as a second factor of authentication," *Communications of the ACM*, vol. 63, no. 12, pp. 46-52, 2020.
- [32] R. A. Grimes, "One-Time Password Attacks," in *Hacking Multifactor Authentication*, Wiley, 2020.

- [33] E. Ulqinaku, D. Lain and S. Capkun, "2FA-PP: 2nd factor phishing prevention," in Proceedings of the 12th Conference on Security and Privacy in Wireless and Mobile Networks, Miami, Florida, 2019.
- [34] O. Umoren and H. Marco-Gisbert, "A Study on the Security of Authentication Systems," in The Fourteenth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies, 2021.
- [35] S. Maaz, G. Sinha and D. K. Sinha, "Examination of Different Network Security Monitoring Tools," in Mobile Computing and Sustainable Informatics. Lecture Notes on Data Engineering and Communications Technologies, Singapore, 2023.

# Enhanced Brain Tumor Detection and Classification in MRI Scans using Convolutional Neural Networks

Ruqsar Zaitoon, Hussain Syed

School of Computer Science Engineering, VIT- AP, Amaravati, Near AP Secretariat, Andhra Pradesh, India

**Abstract**—Tumor detection is one of the most critical and challenging tasks in the realm of medical image processing due to the risk of incorrect prediction and diagnosis when using human-aided categorization for cancer cell identification. Data input is an intensive process, particularly when dealing with a low-quality scan image, due to the background, contrast, noise, texture, and volume of data; when there are many input images to analyze, the task becomes more onerous. It is difficult to distinguish tumor areas from raw MRI scans because tumors pose a diverse appearance and superficially resemble normal tissues, which makes it more difficult to detect tumors. Deep learning techniques are applied in medical images to a great extent to understand tumor contours and areas with high intensities in input images. For timely diagnosis and the right treatment with less human involvement, and to interpret and enhance detection and classification accuracies this automated method is proposed. This proposed work is to identify and classify tumors on 2D MRI scans of the brain. In this work, a dataset is used, inside it, there are images with and without tumors of varied sizes, locations, and forms, with different image intensities and textures. In this paper, multi-layer Convolutional Neural Network (CNN) architectures are implemented. This shows two main experiments to assess the accuracy and performance of the model. First, five-layer CNN architecture with five layers and two different split ratios. Second, six-layer CNN architecture with two different split ratios. In addition, image pre-processing and hyper-parameter tuning were performed to improve the classification accuracy. The results show that the five-layer CNN architecture outperforms the six-layer CNN architecture. When results are compared with state-of-the-art methods, the proposed model for segmentation and classification is better because this model achieved an accuracy of 99.87 percent.

**Keywords**—Multi-layer Convolutional Neural Networks (CNNs); MRI images; tumor segmentation and classification; deep learning; learning rate

## I. INTRODUCTION

Medical imaging is often used by doctors to get a better look at what's going on within a patient's body and arrive at a correct diagnosis. The classification of medical images is not only a formidable intellectual challenge but also a potentially fruitful research field in the field of image processing. Cancers in medical pictures may be difficult to detect. A difficult diagnostic problem has arisen. The importance of cancer screenings cannot be overstated. Death rates and the prevalence of brain tumors attest to the fact that cancer, whether it manifests externally or inside, is a devastating disease. Every year, doctors diagnose more than a million people with tumors. The death toll keeps climbing. In terms of

cancer-related mortality in those under the age of 34, it is second only to lung cancer [1]. In their quest to locate the tumor, doctors are now adopting cutting-edge methods that only serve to make patients more uncomfortable. Computerized tomography scans (CT scans) are used to examine the human body and look for anomalies. Medical Imagination Reasoning (MRI) and the alternatives each have their advantages. There is a rising interest in the study of brain tumors, and the field of image analysis has attracted a lot of attention as a means of analyzing the vast amounts of data available in medical databases. Tools to produce visual representations and complex computational measurements are both necessary for the study of such a wide range of image types. Because of this, MRI scans may now be used to detect malignant brain tumors. This is where the value of handwriting data really shines since it greatly reduces the quantity of paper that would otherwise need to be used. Medical imaging is mostly used for therapeutic and diagnostic purposes in the human body. Thus, it contributes significantly to the development of healthcare and to the betterment of people's lives. In order, to improve the efficiency of image processing as a whole [2], the process of segmentation is vital. This is because it enables the breakdown of the image into its individual elements. We have been diligently working to isolate the tumor in the patient's brain MRIs, providing assistance to medical professionals in pinpointing the precise location of the tumor in the brain. Help is given to medical professionals in pinpointing the specific site of the tumor in the brain. Diagnosis, therapy (including surgical planning), and research all depend on this kind of careful dissection and interpretation. Radiologists, engineers, and doctors all employ medical image processing to learn more about a patient's or a population's unique anatomy. It is possible to get insight into, say, how a patient's anatomy interacts with a medical device via the use of measurement, statistical analysis, and the development of simulation models that contain genuine anatomical geometries. Neoplasm, or tumor, is the medical term for a mass of abnormally growing cells. Cancer and tumor are very different concepts [2]. There are two main subtypes that might each make up a brain tumor. In contrast to malignant tumors, which do contain cancerous cells, benign tumors do not, and vice versa.

1) *Benign tumor*: Benign brain tumors are caused by a disruption in the normal processes of cell division and proliferation, which results in a collection of cells that, on a microscopic level, do not exhibit the classic characteristics of cancer. These traits distinguish benign tumors from malignant ones: Imaging techniques such as computed



tomography (CT) and magnetic resonance imaging (MRI) can identify the great majority of tumors, even benign ones. All these traits point to the fact that this tumor is benign since it develops at a modest rate, and seldom spreads to other parts of the body, which might ultimately result in death, the term "benign" may give the wrong impression about the nature of these injuries.

2) *Malignant tumor*: Development of Carcinoma Cancer cells is the building blocks of malignant brain tumors, which often do not have well-defined borders. The rapid development of these tumors and their ability to invade adjacent brain tissue [4] has led experts to the grim conclusion that they provide a significant risk of death. The following is a list of traits that malignant tumors have malignancy that is quickly spreading, with broad metastases in both spinal and cerebral. Malignant brain tumors are graded as either 3 or 4, whereas benign brain tumors are often categorized as 1 or 2. They usually pose a far bigger danger to human life.

Recent and reliable forecasts [5] estimate that 24,530 persons in the United States will be diagnosed with brain or spinal cord tumors in 2021. There will be 13840 men and 10690 women impacted by this. Less than one percent of the population is expected to get this kind of brain tumor at some point in their life. Roughly 85–90% of primary CNS cancers have this etiology as their root cause. This article focuses on the most common types of brain tumors in adults; however, each year 3,460 children under the age of 15 are diagnosed with a tumor in their central nervous system (CNS). Both men and women rank brain and central nervous system cancer as the ninth biggest cause of death. About 18,600 fatalities in 2021 are projected to be the result of primary brain and central nervous system tumors [6]. There were around 10,500 men and 8,100 women. Therefore, it is crucial to enhance the precision of previously proposed approaches for the advancement of medical image analysis. To summarize, several advanced methods have been introduced to address the problem associated with low image quality. Various techniques have been observed to possess durable effectiveness in improving contrast, illustrating texture intricacies, and reducing noise levels. Moreover, these techniques excessively magnify the intricate features of the images. As per the existing literature paradigms, as far as the Author's knowledge, the issue of brain tumor classification for low-quality MRI scans is yet to be set up to consider it in real-time applications. Hence, to achieve this goal in the present study, a novel technique has been proposed.

The main contribution of this research is summarized as follows:

- A robust multi-layer CNN-based system is proposed for binary-class brain tumor classification on the publicly available dataset.
- An analysis is performed on MRI data because it is one of the main sources for detecting tumors in the patient's brain to detect it.
- The repetition of this invasive method in the case of non-clear images could be avoided if the system is auto

trained with deep learning techniques. To overcome such problems in medical imaging. Hence, an MRI as a data input.

- An enhanced CNN classification model is implemented to identify and classify the brain tumor and compare the achieved accuracy and performance with the state-of-the-art approaches.

The existing framework is offered in the Section II of the paper, the background details are explained in the Section III, the suggested approach is explained in the Section IV, experimental findings and comments are presented in the Section V of the paper, and lastly, a conclusion is presented in the Section VI of the paper.

## II. LITERATURE REVIEW

Khan et al. [7] utilized contrast to lesion area compared to the background. The 2D blue channel is selected for the construction of saliency map, at the end of which threshold function produces the binary image. In addition, particle swarm optimization (PSO) based segmentation is also utilized for accurate border detection and refinement. Few selected features including shape, texture, local, and global are also extracted which are later selected based on genetic algorithm for identifying the fittest chromosome. Hussain et al. [8] devised a way to identify and measure brain tumors using MRI images. The algorithm can identify and eliminate any size, location, or form of tumor. MRI pictures are grayscale first. The blurred picture is then merged with the original. Median filters reduce noise. Dilation and erosion compute morphological gradients. A picture is improved with a morphological gradient and filter. Mean and standard deviation are used to compute the threshold. Before binarization, each pixel's threshold is checked. The image is thinned, and then dilated to reattach the destroyed tumor. Comparing original and dilated photos helps eliminate Javed et al. [9] focus will be to review the optimal features which have been used in an accurate skin cancer melanoma diagnosis computer-aided system. They addressed this problem an extensive review is performed. To perform this review, they collected quality papers based on features selection and extraction for skin lesion detection. These papers are collected by two approached: (1) search by keywords and year, (2) cross-references within the papers.

Javed et al. [10] proposed to deal with under/over segmented images is proposed a region-based active contour method and low contrast skin lesion dermoscopic images handle by implementing JSEG technique. An image fusion technique is proposed on two segmented images get by apply region-based active contour and JSEG techniques. Rashid et al. [11] used a joint design that fuses both the RBAC and JSEG method for skin lesion segmentation. Design technique improved the lesion segmentation as well as deal with the failure cases. The outcomes exhibited an incredible potential by beating state-of-the-art strategies for skin lesion segmentation from thermoscopic images. The proposed method also deals with different artifacts present in the thermoscopic images. They proposed for approach low contrast images by using histograms. Ullah et al [12] proposed early detection and classification of EXs in color fundus

images. An ensemble classification of exudates in color fundus images using an evolutionary algorithm based optimal features selection. Experiment performed on benchmark datasets and a real dataset developed at local Hospital. It has been observed that the proposed technique achieved an accuracy of 98% in the detection and classification of EXs in color fundus images. Sajjad et al. [13] proposed the data augmentation with various parameters and techniques to fill the gap of data and make the system noise invariant. Multi-grade brain tumor classification system, the tumor regions from the dataset are segmented through a CNN model, the segmented data is further augmented using parameters to increase the number of data samples, and a pre-trained VGG-19 CNN model is fine-tuned for multi-grade brain tumor classification. Rehman et al. [14] proposed skull masking method to identify issues. Unsupervised SVMs build and preserve patterns using this approach.

Ahmed et al. [15], which enhances calculation time. The suggested solution hasn't been tested yet. It has 86% classification and 92% cancer detection accuracy. Histograms were utilized by Liu et al. [16] segmenting brain tumors involves two modalities: FLAIR and T1. FLAIR aberrant areas were found using an active contour model. Edema and tumor tissues in aberrant locations were separated using k-means. The dice coefficient is 73.6% and the sensitivity is 90%. Nikam et al. [17] employed edge detection and adaptive thresholding to extract ROI. By using edge detection, the dataset included 102 pictures. First, images were preprocessed, and then two neural network sets underwent canny edge detection and adaptive thresholding. Two neural networks determine whether the brain is healthy or has tumors and the kind of tumor. Canny edge detection was more accurate based on the data and models. Ye et al. [18] enhanced texture-based tumor segmentation in longitudinal MRI using tumor growth patterns. This technique exploits tumor characteristics. Mean DSC LOO and three-Folder measured the model's performance. Sarkar et al. [19] presented a PNN-based LQ model. 18 MRI scans were used for testing and the remainders were used for training. Gaussian filter smoothed photos. Improved PNN cut processing time by 79%. Sharif et al [20] used probabilistic neural networks for segmentation. PCA identified characteristics and reduced high-dimensional data. Mehrotra et al. [[21] proposed a neural network for classifying MRI matrix data and conducted a comprehensive performance analysis with the assistance of deformable models and fuzzy clustering. Rehman et al. [22] separated tumors using Linknet. All seven training datasets were initially segmented in a single Link net network. They didn't examine the perspective of the images and instead created a way for CNN to automatically separate prevalent brain tumors.

Tufail et al. [23] different DL methods are used to solve both binary and multiclass classification problems to differentiate between different stages and deployed a ten-fold cross-validation approach to select the optimal set of hyperparameters for the binary and multiclass classification tasks. For the binary classification task, the performance of architecture trained using combined augmentation methods is the best while the performance of the model trained without

any augmentation is found to be the worst. Other retinal diseases such as retinal detachment using fundus images deploying data augmentation methods such as elastic/plastic deformations as well as other DL-based architectures such as graph convolutional networks.

Pitchai et al. [24] automated deep learning-based Fuzzy K-means clustering segmentation approach has been developed for brain tumor segmentation. This method includes four stages. Initially, the MRI images are preprocessed using a wiener filter for noise expulsion. From the filtered images, the significant features are extricated by using the CSOA algorithm. Then, the normal and abnormal images are classified through ANN. Finally, the fuzzy K-means algorithm has been utilized on the abnormal images to segment the tumor region. This can replace conventional invasive brain tumor classification and enhances the overall classification accuracy. An efficient strategy is employed to enhance the low visual quality of MRI images. Data augmentation technique is used to achieve high classification accuracy on a small dataset, and the impact of over-fitting on classification performance is studied. An efficient and simpler object (tumor) localization method is developed, which gets the initial locations by computing multiple hierarchical segmentation using superpixels and then rank the locations according to region score, which is defined as a number of contours wholly enclosed in the located region, only the top object locations are passed for the next task. Guan et al. [25] proposed a deep neural network (EfficientNet) is employed for rich features extraction. A comparison of the proposed method with existing state-of-the-art approaches for brain tumor classification is presented. The proposed method achieved classification accuracy compared to traditional methods.

Kaplan et al. [[26] proposed two different feature extraction approaches were used to classify the most common brain tumor types; Glioma, Meningiomas, and Pituitary brain tumors; nLBP and  $\alpha$ LBP. Brain tumor classification using modified local binary patterns - nLBP and  $\alpha$ LBP feature extraction methods used. This work introduces an optimized deep learning mechanism; named Dolphin-SCA based Deep CNN, to improve the accuracy and to make effective decisions in classification of brain tumor classification. Kumar et al. [27] presented mechanism of deep learning; named Dolphin-SCA based Deep CNN, to improve the accuracy and to make effective decisions in classification. The segmentation process is carried out using a fuzzy deformable fusion model with Dolphin echolocation-based Sine Cosine Algorithm (Dolphin-SCA). It looks at the posterior and anterior (PA) views of X-rays, therefore it can't tell the difference between other X-ray perspectives like anteroposterior (AP), lateral, and so on. It also needs Grad-CAM (Class Activation Mapping) visualization.

Deepak et al. [28] proposed a classification system that adopts the concept of deep transfer learning and uses a pre-trained GoogLeNet to extract features from brain MRI images. Proven classifier models are integrated to classify the extracted features. The experiment follows a patient-level five-fold cross-validation process, on an MRI dataset from Figshare. They proposed system records a mean classification accuracy of 98%, outperforming all state-of-the-art methods.

Raja et al. [29] developed a brain tumor classification using a hybrid deep autoencoder with a Bayesian fuzzy clustering-based segmentation approach. Initially, the pre-processing stage is performed using the non-local mean filter for denoising purposes. Then the BFC (Bayesian fuzzy clustering) approach is utilized for the segmentation of brain tumors.

Rammurthy et al. [30] presented a fully automated deep CNN for brain tumor detection using MR images. It also proposed a Whale Harris Hawks optimization (WHHO) is employed for training the deep CNN. The proposed WHHO algorithm is designed by combining the WOA and HHO algorithms, which can be utilized for finding the optimal weights for establishing effective brain tumor detection. Here, the segmentation is performed on each input brain MRI image using cellular automata and rough set theory. The pertinent pixel of tumor regions helps to provide improved segmentation results. Advanced optimization techniques can be explored to further compute the efficiency of existing methods.

Agarwal et al. [31] proposed a new Conv2D model for cucumber disease classification with an accuracy of 93.75%, which outperforms the state-of-the-art accuracy by 8.05%. Modified the ReLU activation function and experimentally established that it boosts the classification accuracy by 2.5% over the regular ReLU function. They proposed a segmentation algorithm to identify the diseased regions of cucumber leaf images and work out the severity of the disease. Botta et al. [32], this approach focuses only on relevant image patches from the image, making the technique fast and memory efficient. The low FPR values obtained indicate a low chance of an intact egg being classified as cracked. Thakur et al. [33] proposed to detect and classify Covid-19 disease from normal (healthy) and pneumonia patients; to check the generalizability of the method, develop a larger dataset that includes both X-rays and CTs; to achieve a higher value of performance measures for both binary as well as multiclass problems; to calculate all the performance measures and compares the different parameters of the proposed technique to those of current techniques. Despite having a great performance, it has some drawbacks. In Ullah et al. [34], a medical decision support system using malignant and benignant classes was discussed. This system is designed by median filter, CLAHE, wavelet transform, color moments and feed-forward NN. The proposed system provides results in categorizing malignant and benign MRI images.

It has been noticed from the above research that findings with the traditional machine learning (ML) techniques are not sufficient to segment and classify the brain tumor from raw MRI images. In addition, deep learning techniques outperform the accuracy of brain tumor classification effectively. However, the size of the patient's brain tumor changes periodically, which makes it hard and time-consuming to diagnose and categorize the tumor from massive imaging sets. The structural complexity of the brain makes the building of an expert system for identifying brain tumors a challenging undertaking that is plagued by several problems, including under-fitting, biased results, overfitting, and repetition of the training samples. It takes more time to complete activities like determining the infected area, segmenting, and identifying

tumors from MRIs, and it is challenging to see the abnormal brain structures using standard image processing methods. The present proposal clearly put a light on the Convolutional neural network which is the backbone of every other deep neural network. Thus, authors have tried to show the working of CNN with a slight change in layers and proportions to advance the accuracy and performance to overcome the problem of accurate detection and classification of brain tumours in MRI scans (see Table I).

TABLE I. OVERVIEW OF LITERATURE REVIEW

Reference	Objective	Methods use	Outcome
Hussain et al. [8]	Identify and measure brain tumors in MRI images.	Median filters, morphological gradients, thresholding, image processing.	Noise reduction, tumor identification, size measurement.
Javed et al. [9]	Detect brain tumor malignancy.	PCA, RST, wavelets, CNN, bilateral filters, histogram equalization.	Noise reduction, feature extraction, CNN classification.
Rahim et al. [10]	Recognize tumor blocks and types.	High-pass filters, K-means clustering, neural networks.	Segmentation, classification, noise reduction.
Rashid [11]	Employ morphological approaches and filtering.	Pixel removal, thresholding, skull masking.	Tumor segmentation, pattern recognition, noise reduction.
Ahmed et al. [15]	Improve calculation time, classification accuracy.	Not specified.	Classification accuracy improvement, untested solution.
Liu et al. [16]	Segment brain tumors using FLAIR and T1 modalities.	Active contour model, k-means, edge detection, neural networks.	Aberrant area detection, sensitivity, dice coefficient.
Nikam et al. [17]	Extract ROI and classify brain tumors.	Edge detection, adaptive thresholding, neural networks.	Edge detection accuracy, tumor classification.
Ye et al. [18]	Enhance texture-based tumor segmentation.	Texture-based features, mean DSC LOO, 3-Folder, longitudinal MRI.	Texture-based segmentation, performance measurements.
Sarkar et al. [19]	Present a PNN-based LQ model.	Gaussian filter, probabilistic neural networks, PCA.	Processing time reduction, feature identification.
Sharif et al. [20]	Use probabilistic neural networks for segmentation.	PCA, neural network classification.	High-dimensional data reduction, performance analysis.
Rehman et al. [22]	Separate tumors using Linknet.	Linknet network, CNN, prevalent brain tumors.	Automatic tumor separation, CNN-based approach.
Guan et al. [25]	Improve visual quality, tumor location proposals.	Efficient-net, image-pre-processing.	Tumor segmentation in multi-modal images.
Kaplan et al. [26]	Classify common brain tumor types.	Local binary patterns, nLBP, $\alpha$ LBP.	Tumor type classification based on feature extraction.

Rammurthy et al [30]	Optimization employed to obtain optimized weight.	Whale Harris Hawks Optimization algorithm used	Improved segmentation, and accuracy.
----------------------	---	--	--------------------------------------

### III. BACKGROUND

#### A. Convolutional Neural Network

CNN is an example of an artificial neural network, often known as a multilayered sensor. Its architecture is inspired on the structure of the visual cortex. One of the most important ideas behind deep learning is the Convolutional Neural Network or CNN. CNN consists of two basic processes, which are referred to as convolution and pooling, and is often used in applications that are related to image recognition. Additional layers of convolution and pooling are added as necessary up to the point when a high level of classification accuracy is obtained. In addition to this, certain feature maps are included in each convolutional layer, and the weights of convolutional nodes that are contained inside the same map are shared. These designs minimize the number of traceable parameters while allowing for the learning of a variety of network properties. CNN, in contrast to more traditional methods, may learn to completely extract features while simultaneously performing a reduced number of specialized duties. The whole process plan of a CNN is shown in Fig. 1.

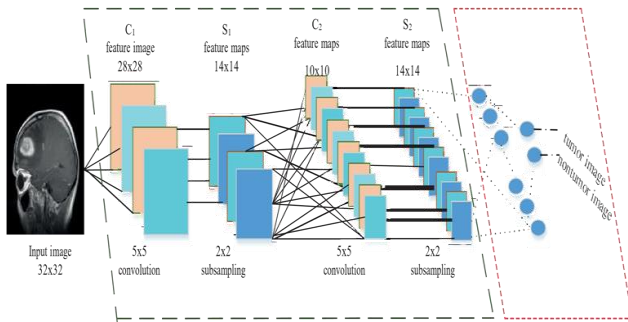


Fig. 1. CNN process.

#### B. Background of Tumor Detection

Preprocessing: Noise may affect MR images. Picture compression and data transfer may produce noise. Nonlocal techniques and local smoothing were used to reduce noise [ ]. Some significant visual structures and features might appear as if they were made of noise; these vital details may also be deleted. Below figure shows axial, coronal, and sagittal MR images. Fig. 2(a) and 2(b) show an original and preprocessed image.

#### C. Watershed Segmentation and Morphological Process

Good cranial MRI segmentation using watershed segmentation. This method detected cancer. Geography and water supply basins determine watershed lines. It analyses grey data and determines the object's boundaries by using the topological structure of the object inside an image. It enhanced the submerge-based watershed transformation method. It's lowest and maximum values are hmin and hmax. Moving from hmin to hmax shows recursion. Xh basin clusters were comparable to dot clusters with hmin at recursion's start. The Xh basin cluster in the threshold cluster eventually grows.

$$Xh = \min_U IZTh + 1(f)(Xh), \forall h \in [hmin, hmax - 1] \quad (1)$$

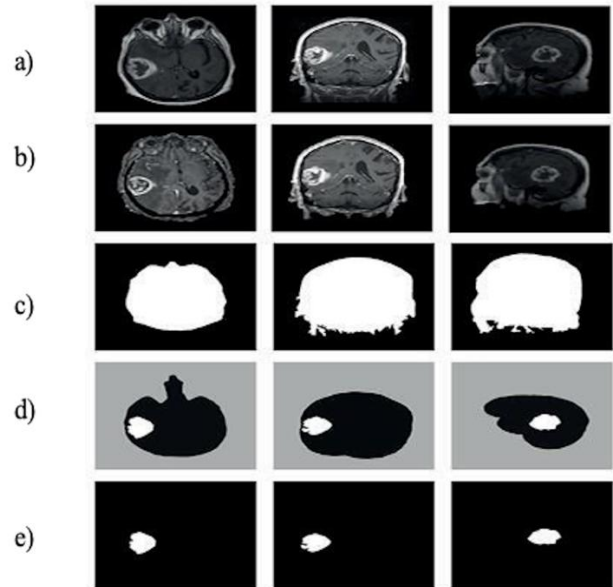


Fig. 2. (a) Original image, (b) Pre-processed image, (c) Segmented skull image, (d) Brain tissue extraction, (e) Tumor detection.

First, the program collects gradient data. Subtracting the first derivative of pixel change yields this data. Next-level activation requires the signal. Image segmentation requires pointer pixels for each class. These pixels' location and quantity affect segmentation success. The watershed transformation has interesting features for mathematical morphology image segmentation. This change is intuitive. It generates closed curves quickly. Over-segmentation is a nearby one fused, and gaps were filled. Fig 2(c) shows how watershed segmentation eliminates data from MR images. Fig. 2d shows how the brain's soft tissue was created by removing the skull from the original image. Fig 2(e) shows how morphological segmentation and classification of brain tumour in MRI scans and assist the radiologist.

### IV. PROPOSED WORK

The proposed five-layer CNN model can identify a tumor in MRI images. Fig. 3 shows the five-layer CNN technique. Firstly, load the input dataset with the same-sized images. Five-layer CNN is used for early identification of tumors and a model includes seven steps (eight if you include the hidden layers), yielding the best results for tumor detection and splitting down the process into seven steps from which a brain tumor may be detected utilizing CNN. Step-by-step instructions shown below reveal the process. In Fig. 4, the proposed method for tumor detection using a five-layer Convolutional Neural Network is shown. Five discrete dimensions are also used.

#### A. Convolutional Layer

A convolutional layer is CNN's backbone. First, a convolutional layer is used to resize MRI images. This makes a 64\*64\*3 input shape. After gathering all same-orientation input images developed a convolutional kernel using 32 3\*3 convolutional filters and three channel tensors. The activation

function is ReLU. The filter size is three times the 64x64x3 input volume.

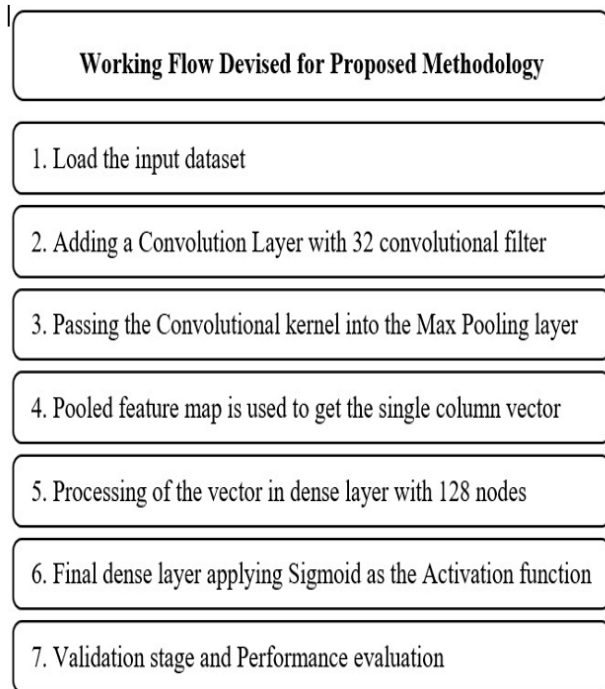


Fig. 3. Five layer CNN model workflow.

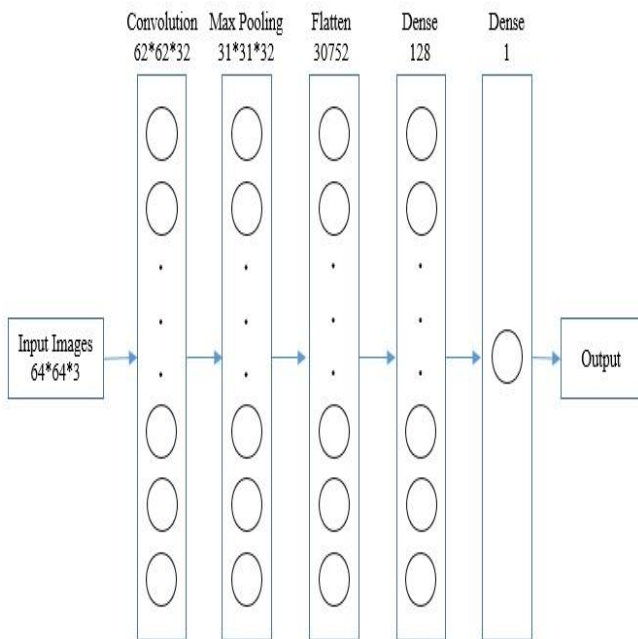


Fig. 4. Five-layer CNN for brain tumour detection.

Each neuron in the convolutional layer has  $3*3*3 = 27$  weights, plus one for the bias parameter. Assess depth, stride, and zero-padding. The model contains a  $64*64*3$  input volume and a 33 spatial filter. Since it didn't specify border padding, padding and stride are both 1. If the stride is set to 1,

only the pooling layers will down sample; the CONV layers will alter the input volume in depth. After the convolutional layer, added max pooling.

**B. Max Pooling Layer**

The fundamental objective of the pooling layer is to reduce the number of parameters and computation workloads in the network by progressively decreasing the spatial size of the representation. Over-fitting may be managed thanks to its ability to scale down the settings. The max pooling layer may be used to enlarge the input spatially, and it can do so on a per-slice basis if the input is deep enough. From another angle, the Max Pooling layer is great for preventing over-fitting, which might introduce contamination into the brain MRI image when editing it (see Fig. 5). Therefore, MaxPooling2D (see Fig. 6) check the effect of the pooling operation as a pre-processing step which was used on the input image. There are a total of 32 nodes in this convolutional layer, resulting in a  $31*31*31$  matrix.

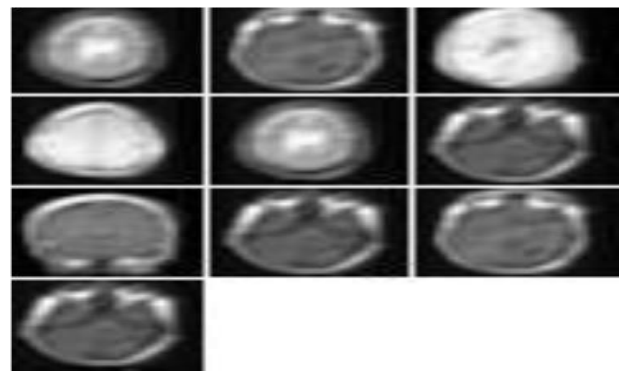


Fig. 5. CNN operation unit.

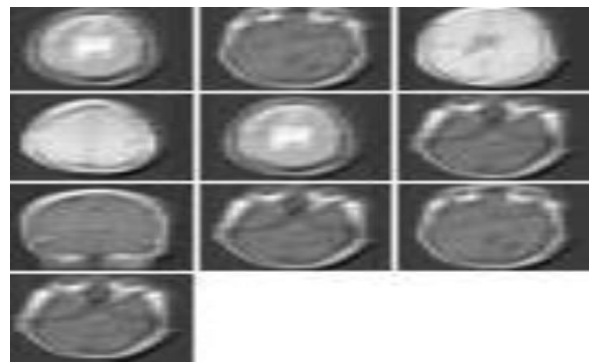


Fig. 6. Pooling operation unit.

**C. Flatten Layer**

Pooling creates a pooled feature map that extracts necessary features only and discards unimportant features refer to Fig. 6. The flattened layer is crucial after pooling because it transforms the input's feature maps into a single-column vector for processing. The neural network processes it as Layer  $31*31*32 = 30752$  pixels.

**D. Fully Connected Layer**

The dense-1 and dense-2 (see Fig. 7) layers were linked. Keras processes the neural network using the dense function, and the output vector is fed to this layer. Each layer has 128

nodes. Due to high processing costs, the number of dimensions or nodes is 128. ReLU is employed due to its high convergence. The model's final layer was the second totally linked layer, utilizing the sigmoid function as the activation function in this layer with one node to reduce execution time. Sigmoid activation may impair deep network learning. The sigmoid function has been lowered, reducing the number of nodes in this deep network.

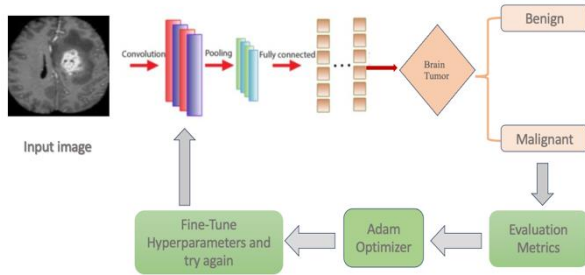


Fig. 7. Proposed model workflow.

## V. RESULTS AND DISCUSSION

### A. Experimental Setup

Jupyter Notebook and Python technologies like NumPy, Pandas, and OpenCV were used for image processing. For classifiers, use Scikit-Learn, Anaconda and Python 3.6.

CNN model was trained and tested using TensorFlow and Keras used Google Colab's GPU.

### B. Dataset Acquisition

Br35H-Mask-RCNN dataset [16] is used to segment brain tumors. Cancers and non-tumors are labeled. A set has two classes. Class-1 is tumors MRI while Class-2 is non-tumors (class-0). Training is 700, testing is 100 and 24 images test performance.

### C. Performance Measures

To assess how well the proposed model works, one must consider performance metrics. Below is the discussion on the proposed model's performance statistics. One must learn performance measurement lingo to understand the Confusion Matrix.

Confusion Matrix:

TP: Number of correctly identified tumor images.

TN: The number of correctly detected non-tumor images.

FP: The number of non-tumor images labeled tumor.

FN: Number of tumor images misclassify as non- tumor.

Accuracy: It's the most common way to measure how often a classifier is correct. Accuracy is the ratio of accurately predicted images to the total number of photos.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

Precision: The model's data is retrieved. Precision is the ratio of correctly detected tumor images (TP) to misclassified

ones (TP+FP). Precision rises with FP. A more accurate model is more effective. It's the proportion of retrieved images.

$$Precision = \frac{TP+TN}{TP+FP} \quad (3)$$

Recall: Recall is the ratio of tumor photographs correctly detected to images to be projected. It's also called sensitivity, hit rate, and true positive rate. Because non-tumor images are rare, a smaller false negative increases memory.

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

F-Score: The harmonic mean of recall and precision is used as a measurement of test accuracy. F-scores range from 1 (perfect accuracy and recall) to 0 (zero).

$$F - Score = \frac{2TP}{2TP+FP+FN} \quad (5)$$

Specificity: Specificity is the F-Score. The model's True Negative Rate (TNR) and binary classification test's statistical measure is specificity. Use this as a performance evaluation as it's binary (tumor or non-tumor). The ratio of correctly identified non-tumor images (TN) to wrongly categorized non-tumor images (TN + FP). The more specificity, the fewer false positives (FP).

$$Specificity = \frac{TN}{TN+FP} \quad (6)$$

### D. Experimental Results

The five-layer CNN model got the best results for splitting ratio and other parameters. Then split CNN model performance by layer count. Next, is presented the experimental results and quality and evaluation. Experiments I and II tested the five-layer model with 70:30 and 80:20 learning rates and epoch-splitting ratios. Later, it is examined as five, six, and seven-layer CNN models.

1) *Experiment-I*: The five-layer CNN model is trained using 70 by 30 ratios. Table II shows how this ratio impacts learning rate, epochs, training length, and accuracy. The highest accuracy was attained using a 0.001 learning rate, 50 epochs, and 500 seconds of training.

TABLE II. CNN TRAINING TIME ACCURACY (SPLITTING RATIO OF 80:20)

Learning Rate	Time	Time to train(sec)	Accuracy (%)
0.001	10	175	99.51
	20	233	98.87
	50	527	95.74
	100	1200	95.69
0.005	10	177	96.03
	20	203	97.62
	50	488	95.55
	100	1027	95.55
0.01	10	178	92.09
	20	200	93.04
	50	599	93.77
	100	966	92.00

2) *Experiment – II:* The five-layer CNN model is trained using 80 by 20 ratios. Table III compares training, accuracy, learning rate, and epochs. The greatest accuracy is 99.51 percent at a 0.001 learning rate and 10 epochs, and training takes 175 seconds.

TABLE III. CNN TRAINING TIME ACCURACY (SPLITTING RATIO OF 70:30)

Learning Rate	Time	Time to train(sec)	Accuracy (%)
0.001	10	180	98.66
	20	231	98.95
	50	500	99.58
	100	1227	96.01
0.005	10	198	98.18
	20	240	99.13
	50	555	97.68
	100	1133	97.22

3) *Experiment-III (Five-layer Architecture):* This analysis used several hyper-parameters and splitting ratios to test the five-layer CNN model. Table IV compares 80:20 and 70:30 five-layer CNNs. The model is most accurate for 80:20, 64, and 10 epochs. The model overfits thereafter. Five-layer CNN accuracy is 99.87%.

TABLE IV. CNN PERFORMANCE FOR FIVE LAYERS

Convolution Layer	Coalescing	Fracture Percentage	Group Measurement	Time	Accuracy (%)
62*62*32 62*62*32 (2 layer)	31*31*32	80:20	32	8	89.29
				9	92.83
				10	93.76
				11	93.62
			64	8	94.01
				9	96.08
				10	95.49
				11	94.21
		70:30	32	8	82.37
				9	83.71
				10	84.24
				11	86.17
			64	8	88.27
				9	82.23
				10	81.69
				11	80.07

4) *Experiment-IV (Six-layer Architecture):* Add 62\*62\*32 convolutional layer. Changing model dimensions may affect accuracy. Table V demonstrates 80:20 splitting, 64 batches, and 11 epochs. Model accuracy deteriorated. Convolutional layers don't improve accuracy. Two kinds of CNN models

were employed for the suggested model, and the five-layer CNN model had 99.87% accuracy. Table VI compares an CNN models from trial III and IV.

TABLE V. CNN PERFORMANCE FOR SIX LAYERS

Convolution Layer	Coalescing	Fracture Percentage	Group Measurements	Time	Accuracy (%)
62*62*32	31*31*32	80:20	32	8	99.87
				9	99.77
				10	99.51
			64	8	93.67
				9	94.98
				10	97.87
				11	94.89
				11	94.90
		70:30	32	8	81.35
				9	83.71
				10	87.87
				11	89.13
			64	8	88.07
				9	88.76
				10	91.23
				11	94.90

TABLE VI. COMPARISON OF CNN MODELS

No. of Layers	Convolutional Layer	Coalescing	Fracture Percentage	Group Measurements	Time	Accuracy (%)
5	62*62*32	31*31*32	80:20	64	8	99.87
6				64	9	96.08

*E. Discussion*

The proposed method uses a pre-processing, data augmentation, convolutional kernel, filter, and three channel tensors added in the process to improve classification. Techniques like MaxPooling2D were used to reduce the overfitting problem and added two fully connected dense layers with ReLU as an activation function.

A detailed discussion on the effect of execution time, variation in proposed results, and comparison with different add or removal of layers in CNN are projected here. The process is implemented on two different architecture styles and measures the accuracy and system computational time, such as 80:20, and 70:30 on five-layer CNN architecture which can be viewed from Table II and Table V. In these tables, it is shown that the best-achieved accuracy is 99.87%, 99.77%, 99.51% etc. The variant performance of the proposed methodology is noticed when we add or drop the classifier layers and training proportions on the proposed model. The proposed selected features are best for five-layer CNN. Also, applied hyperparameter tuning on six-layer CNN. However, the results are not up to the scale (see Table V) when

compared with five-layer CNN, 80:20 training split ratio (see Table IV) and observed that this method improves the computational and classification accuracy.

List of Hyperparameters used in this study are Batch size, Learning Rate, number of epochs, number of filters, kernel size, pooling size. Refer to Table II and Table III to check tuning performance of proposed model. In addition, Adam optimizer is used. The above parameters can fine-tune each time we find less accuracy as per the evaluation metrics see Fig. 7. To evaluate the proposed model plotted confusion matrix (See Fig. 8)

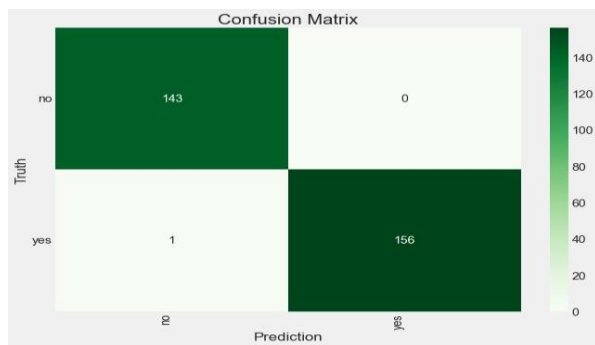


Fig. 8. Confusion matrix.

## VI. CONCLUSION AND FUTURE WORKS

In this analysis, the enhanced tumor segmentation and classification from MR images is presented. The proposed model used two techniques with CNN which works in two phases: (1) five-Layer CNN with Hyper-parameter tuning; (2) six-layer CNN with hyperparameter tuning with various data training proportions. This model performed well because of 2D Max pooling, ReLu activation function and used Adam as optimizer and fine tune with batch size, learning rate. In a proposed approach, the best features selection step not only increases the accuracy but also minimizes the classification time. However, this study has limitation for computing large datasets and complex CNN's this model may not give same accuracy to other datasets. In the future, we aim to extend our current work for fine-grained classification of multi-modal MRI images on advanced CNN architectures like R-UNet and aim to use a high-performance computational platform for the complete dataset for further improvement in segmentation feature map and classification. This soon can be deployed in real-time applications in a broader prospect.

## REFERENCES

- [1] Amin, J., Sharif, M., Raza, M., Saba, T., & Anjum, M. A. (2019). Brain tumor detection using statistical and machine learning method. *Computer methods and programs in biomedicine*, 177, 69-79.
- [2] Amin, J., Sharif, M., Raza, M., Saba, T., & Rehman, A. (2019, April). Brain tumor classification: feature fusion. In *2019 international conference on computer and information sciences (ICCIS)* (pp. 1-6). IEEE.
- [3] Saba, T., Khan, M. A., Rehman, A., & Marie-Sainte, S. L. (2019). Region extraction and classification of skin cancer: A heterogeneous framework of deep CNN features fusion and reduction. *Journal of medical systems*, 43(9), 289.
- [4] Saba, T., Khan, S. U., Islam, N., Abbas, N., Rehman, A., Javaid, N., & Anjum, A. (2019). Cloud-based decision support system for the detection and classification of malignant cells in breast cancer using

- breast cytology images. *Microscopy research and technique*, 82(6), 775-785.
- [5] Khan, M. A., Lali, I. U., Rehman, A., Ishaq, M., Sharif, M., Saba, T., ... & Akram, T. (2019). Brain tumor detection and classification: A framework of marker-based watershed algorithm and multilevel priority features selection. *Microscopy research and technique*, 82(6), 909-922.
- [6] Khan, S. A., Nazir, M., Khan, M. A., Saba, T., Javed, K., Rehman, A., & Awais, M. (2019). Lungs nodule detection framework from computed tomography images using support vector machine. *Microscopy research and technique*, 82(8), 1256-1266.
- [7] Khan, M. A., Akram, T., Sharif, M., Saba, T., Javed, K., Lali, I. U., ... & Rehman, A. (2019). Construction of saliency map and hybrid set of features for efficient segmentation and classification of skin lesion. *Microscopy research and technique*, 82(6), 741-763.
- [8] Khan, M. Q., Hussain, A., Rehman, S. U., Khan, U., Maqsood, M., Mehmood, K., & Khan, M. A. (2019). Classification of melanoma and nevus in digital images for diagnosis of skin cancer. *IEEE Access*, 7, 90132-90144.
- [9] Javed, R., Rahim, M. S. M., Saba, T., & Rehman, A. (2020). A comparative study of features selection for skin lesion detection from dermoscopic images. *Network Modeling Analysis in Health Informatics and Bioinformatics*, 9, 1-13.
- [10] Javed, R., Saba, T., Shafry, M., & Rahim, M. (2019, October). An intelligent saliency segmentation technique and classification of low contrast skin lesion dermoscopic images based on histogram decision. In *2019 12th International Conference on Developments in eSystems Engineering (DeSE)* (pp. 164-169). IEEE.
- [11] Javed, R., Rahim, M. S. M., Saba, T., & Rashid, M. (2019). Region-based active contour JSEG fusion technique for skin lesion segmentation from dermoscopic images. *Biomedical Research*, 30(6), 1-10.
- [12] Ullah, H., Saba, T., Islam, N., Abbas, N., Rehman, A., Mehmood, Z., & Anjum, A. (2019). An ensemble classification of exudates in color fundus images using an evolutionary algorithm based optimal features selection. *Microscopy research and technique*, 82(4), 361-372.
- [13] Sajjad, M., Khan, S., Muhammad, K., Wu, W., Ullah, A., & Baik, S. W. (2019). Multi-grade brain tumor classification using deep CNN with extensive data augmentation. *Journal of computational science*, 30, 174-182.
- [14] Rehman, A., Naz, S., Razzak, M. I., Akram, F., & Imran, M. (2020). A deep learning-based framework for automatic brain tumors classification using transfer learning. *Circuits, Systems, and Signal Processing*, 39, 757-775. [CrossRef] Jacobs, I. S., & Bean, C. P. (1963). Fine particles, thin films and exchange anisotropy (effects of finite dimensions and interfaces on the basic properties of ferromagnets). *Spin arrangements and crystal structure, domains, and micromagnetics*, 3, 271-350.
- [15] Ahmad, I., Ullah, I., Khan, W. U., Ur Rehman, A., Adrees, M. S., Saleem, M. Q., ... & Shafiq, M. (2021). Efficient algorithms for E-healthcare to solve multiobject fuse detection problem. *Journal of Healthcare Engineering*, 2021, 1-16.
- [16] Ahmad, I., Liu, Y., Javeed, D., & Ahmad, S. (2020, May). A decision-making technique for solving order allocation problem using a genetic algorithm. In *IOP Conference Series: Materials Science and Engineering* (Vol. 853, No. 1, p. 012054). IOP Publishing.
- [17] Nikam, R. D., Lee, J., Choi, W., Banerjee, W., Kwak, M., Yadav, M., & Hwang, H. (2021). Ionic Sieving Through One-Atom-Thick 2D Material Enables Analog Nonvolatile Memory for Neuromorphic Computing. *Small*, 17(44), 2103543.
- [18] Ye, F., & Yang, J. (2021). A deep neural network model for speaker identification. *Applied Sciences*, 11(8), 3603.
- [19] Ijaz; Liu; Javeed; Shamshad; Sarwr; Ahmad (19). AI selection and assessment procedures. 2020, 853, 012055. Reconstructing comprehensible speech from the human auditory cortex. 2019;9(8)74.
- [20] Saleem, S., Amin, J., Sharif, M., Anjum, M. A., Iqbal, M., & Wang, S. H. (2021). A deep network designed for segmentation and classification of leukemia using fusion of the transfer learning models. *Complex & Intelligent Systems*, 1-16.
- [21] Mehrotra, R., Ansari, M. A., Agrawal, R., & Anand, R. S. (2020). A transfer learning approach for AI-based classification of brain tumors. *Machine Learning with Applications*, 2, 100003.
- [22] Ramzan, F., Khan, M. U. G., Rehmat, A., Iqbal, S., Saba, T., Rehman, A., & Mehmood, Z. (2020). A deep learning approach for automated diagnosis and multi-class classification of Alzheimer's disease stages using resting-state fMRI and residual neural networks. *Journal of medical systems*, 44, 1-16.



- [23] Tufail, A. B., Ullah, I., Khan, W. U., Asif, M., Ahmad, I., Ma, Y. K., ... & Ali, M. S. (2021). Diagnosis of diabetic retinopathy through retinal fundus images and 3D convolutional neural networks with limited number of samples. *Wireless Communications and Mobile Computing*, 2021, 1-15.
- [24] Pitchai, R., Supraja, P., Victoria, A. H., & Madhavi, M. J. N. P. L. (2021). Brain tumor segmentation using deep learning and fuzzy K-means clustering for magnetic resonance images. *Neural Processing Letters*, 53, 2519-2532.
- [25] Guan, Y., Aamir, M., Rahman, Z., Ali, A., Abro, W. A., Dayo, Z. A., ... & Hu, Z. (2021). A framework for efficient brain tumor classification using MRI images.
- [26] Kaplan, K., Kaya, Y., Kuncan, M., & Ertunç, H. M. (2020). Brain tumor classification using modified local binary patterns (LBP) feature extraction methods. *Medical hypotheses*, 139, 109696.
- [27] Kumar, S., & Mankame, D. P. (2020). Optimization driven deep convolution neural network for brain tumor classification. *Biocybernetics and Biomedical Engineering*, 40(3), 1190-1204.
- [28] Deepak, S., & Ameer, P. M. (2019). Brain tumor classification using deep CNN features via transfer learning. *Computers in biology and medicine*, 111, 103345.
- [29] Raja, P. S. (2020). Brain tumor classification using a hybrid deep autoencoder with Bayesian fuzzy clustering-based segmentation approach. *Biocybernetics and Biomedical Engineering*, 40(1), 440-453.
- [30] Rammurthy, D., & Mahesh, P. K. (2022). Whale Harris hawks optimization based deep learning classifier for brain tumor detection using MRI images. *Journal of King Saud University-Computer and Information Sciences*, 34(6), 3259-3272. [CrossRef]Resize Function. Available online: <https://www.mathworks.com/products/matlab.html> (accessed on 4 January 2022) .
- [31] A Agarwal, M., Gupta, S., & Biswas, K. K. (2021). A new Conv2D model with modified ReLU activation function for identification of disease type and severity in cucumber plant. *Sustainable Computing: Informatics and Systems*, 30, 100473.
- [32] Botta, B., Gattam, S. S. R., & Datta, A. K. (2022). Eggshell crack detection using deep convolutional neural networks. *Journal of Food Engineering*, 315, 110798. [CrossRef]
- [33] Thakur, S., & Kumar, A. (2021). X-ray and CT-scan-based automated detection and classification of covid-19 using convolutional neural networks (CNN). *Biomedical Signal Processing and Control*, 69, 102920.
- [34] Ullah, Z., Farooq, M. U., Lee, S. H., & An, D. (2020). A hybrid image enhancement based brain MRI images classification technique. *Medical hypotheses*, 143,10992

# Method for Hyperparameter Tuning of Image Classification with PyCaret

Kohei Arai<sup>1</sup>, Jin Shimazoe<sup>2</sup>, Mariko Oda<sup>3</sup>

Information Science Dept., Saga University, Saga City, Japan<sup>1</sup>  
Applied AI Laboratory, Kurume Institute Technology, Kurume City, Fukuoka, Japan<sup>1, 2, 3</sup>

**Abstract**—A method for hyperparameter tuning of image classification with PyCaret is proposed. The application example compares 14 classification methods and confirms that Extra Trees Classifier has the best performance among them, AUC=0.978, Recall=0.879, Precision=0.969, F1=0.912, Time=0.609 bottom. The Extra Trees Classifier produces a large number of decision trees, similar to the random forest algorithm, but with random sampling of each tree and no permutation. This creates a dataset for each tree containing unique samples, and from the ensemble set of features a certain number of features are also randomly selected for each tree. The most important and unique property of the Extra Trees Classifier is that the feature split values are chosen randomly. Instead of using Gini or entropy to split the data to compute locally optimal values, the algorithm randomly selects split values. This makes the tree diverse and uncorrelated. i.e. the diversity of each tree. Therefore, it is considered that the classification performance is better than other classification methods. Parameter tuning of Extra Trees Classifier was performed, and training performance, test performance, ROC curve, accuracy rate characteristics, etc. were evaluated.

**Keywords**—PyCaret; extra trees classifier; AUC; gini; entropy; feature split; ROC curve

## I. INTRODUCTION

Most machine learning problems require many hyperparameter tunings. Unfortunately, it is not possible to provide specific tuning rules for all models but may converge very slowly for another model. Finding the best set of hyperparameters for your dataset requires experimentation. Some rules of thumb, here is the training loss, which should decrease abruptly at first, then more slowly and steadily until the slope of the curve reaches or approaches zero, and if the training loss does not converge, wait for more epochs of training.

If the training loss decreases too slowly, increase the learning rate. Note that setting the learning rate too high can prevent the training loss from converging. Decrease the learning rate if the training loss varies a lot (that is, if the training loss jumps around). Increasing the number of epochs or batch size while decreasing the learning rate is often a good combination. Setting the batch size to a very small batch number can lead to instability. First, try increasing the batch size value. Then reduce the batch size until you see a drop. For real datasets consisting of a very large number of samples, the entire dataset may not fit in memory. In such cases, the batch size should be reduced so that the batch fits in memory. For the optimization of these hyperparameters, hyperopt, gpyopt,

AutoML, PyCaret, Optuna, etc. are proposed as black-box optimization methods, automating trial-and-error on hyperparameters and automatically finding the optimal solution. Similarly, a method for training and white boxing DL, BDT, random forest and mind maps based on GNN is proposed [1]. In particular, Optuna uses an algorithm called TPE (Tree-structured Parzen Estimator), which is a new technique among Bayesian optimization, and parallel processing is possible, and by saving the results in a database, it is possible to resume in the middle. Depending on the definition of the objective function and the validity of the importance of the parameters, hyperparameters that do not necessarily match the evaluation criteria may appear. Usually, it is desirable to introduce such cases, and propose a method to intentionally change the hyperparameters of Optuna and PyCaret and select parameters with less loss by trial and error. It, however, PyCaret only uses hyperparameters obtained by random grid search, and does not intentionally change hyperparameters in this paper.

As a comparative study on classification methods with PyCaret and its application to textile fluctuations of classifications, we propose a comparative study of classification methods by PyCaret and its application to the classification evaluation of textile pattern deviations. Many classification methods have been applied to the classification evaluation of textiles, but trial and error are necessary to determine the optimum method, which requires not a little time. The method proposed in this paper finds the optimal method using PyCaret, and is a method for finding the optimal method easily in a relatively short time without repeating trial and error. As one application of this method, an example of applying it to classification of Kurume Kasuri patterns is shown. This is just one application example, and the proposed method can be widely applied to other classifications.

The application example compares 14 classification methods and confirms that Extra Trees Classifier has the best performance among them, AUC=0.978, Recall=0.879, Precision=0.969, F1=0.912, Time=0.609 bottom. The Extra Trees Classifier produces a large number of decision trees, similar to the random forest algorithm, but with random sampling of each tree and no permutation. This creates a dataset for each tree containing unique samples, and from the ensemble set of features a certain number of features are also randomly selected for each tree. The most important and unique property of the Extra Trees Classifier is that the feature split values are chosen randomly. Instead of using Gini or entropy to split the data to compute locally optimal values, the

algorithm randomly selects split values. This makes the tree diverse and uncorrelated i.e. the diversity of each tree. Therefore, it is considered that the classification performance is better than other classification methods. Parameter tuning of Extra Trees Classifier was performed, and training performance, test performance, ROC curve, accuracy rate characteristics, etc. were evaluated.

In the following section, some of the related research works are described together with a research background, followed by the comparative study conducted. Then, some of the simulation studies are described, followed by a conclusion with some discussions.

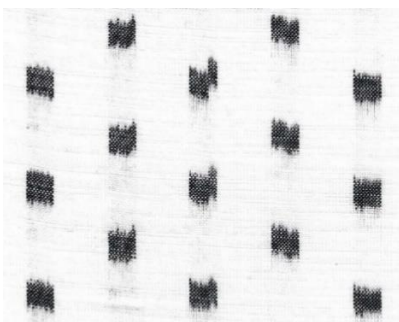
## II. RELATED RESEARCH WORKS AND RESEARCH BACKGROUND

### A. Research Background

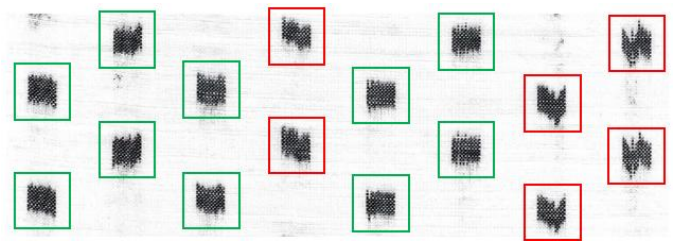
Image classification is required for textile pattern discrimination in this concern. Patterns are boring and uninteresting because they are regular, but moderately irregular patterns feel comfortable. For instance, Kurume Kasuri textile patterns are moderately irregular so that we feel these patterns are comfortable. It is understandable that the moderately irregular patterns contain 1/f fluctuations. However, a pattern with too much irregularity gives an unpleasant feeling to the person who sees it. Therefore, it is thought that there is a boundary between pleasant irregularities and unpleasant irregularities. The purpose of this paper is to discriminate between both. In order to find the best discrimination performance of classification method, 14 of classification methods are compared these accuracies and their other performances.

### B. Examples of Pleasant and Unpleasant Irregularity of Kurume Kasuri

One of the typical Kurume Kasuri of textile patterns is shown in Fig.1 (a). Also, examples of the regular (green) and the moderately irregular patterns (red) are shown in Fig. 1 (b), respectively. As shown in Fig. 1 (b), it is obvious that the regular patterns make a good impression while the irregular patterns make a bad impression. The cause of these irregular patterns is the lack of control over the warp tension of the Kurume Kasuri automatic loom. If the pattern is too comfortable, it will exceed the allowable limit and become an unpleasant pattern. It is too difficult to control the warp tension. It is important to evaluate the quality of woven Kurume Kasuri and determine whether it is good or bad to send it to the market.



(a) Typical Kurume Kasuri of textile pattern



(b) Examples of the regular (green) and the moderately irregular patterns (red)  
Fig. 1. Typical kurume kasuri textile pattern and example of the regular and the moderately irregular patterns.

### C. Related Research Works

Method for 1/f fluctuation component extraction from images and its application to improve Kurume Kasuri quality estimation is proposed [2]. In this paper, it is shown that Kurume Kasuri textile pattern quality depends on 1/f component. On the other hand, classification by re-estimating statistical parameters based on auto-regressive model is proposed [3]. Similarly, multi-temporal texture analysis in TM classification is proposed [4] together with Maximum Likelihood (MLH) TM classification taking into account pixel-to-pixel correlation [5]. Meanwhile, a supervised TM classification with a purification of training samples is proposed [6]. Also, TM classification using local spectral variability is proposed [7]. Furthermore, a classification method with spatial spectral variability is proposed [8].

An inversion for emissivity-temperature separation with ASTER data is proposed [9]. TM classification using local spectral variability is also proposed [10]. On the other hand, application of inversion theory for image analysis and classification is proposed [11]. Meanwhile, polarimetric SAR image classification with maximum curvature of the trajectory in eigen space domain on the polarization signature is proposed [12]. Moreover, a hybrid supervised classification method for multi-dimensional images using color and textural features are proposed [13].

Human gait gender classification using 3D discrete wavelet transformation feature extraction is proposed [14]. Meanwhile, polarimetric SAR image classification with high frequency component derived from wavelet multi resolution analysis: MRA is proposed [15]. On the other hand, a comparative study of polarimetric SAR classification methods including proposed method with maximum curvature of trajectory of backscattering cross section in ellipticity and orientation angle space is proposed [16]. Human gait gender classification using 2D discrete wavelet transforms energy is proposed [17]. Also, human gait gender classification in spatial and temporal reasoning is proposed [18]. Comparative study on discrimination methods for identifying dangerous red tide species based on wavelet utilized classification methods is conducted [19].

Multi spectral image classification method with selection of independent spectral features through correlation analysis is proposed [20]. On the other hand, image retrieval and classification method based on Euclidian distance between normalized features including wavelet descriptor is also proposed [21].

Gender classification method based on gait energy motion derived from silhouettes through wavelet analysis of human gait moving pictures is proposed [22]. Similarly, human gait skeleton model acquired with single side video camera and its application and implementation for gender classification is proposed [23] together with human gait skeleton model acquired with single side video camera and its application and implementation for gender classification [24]. Gender classification method based on gait energy motion derived from silhouette through wavelet analysis of human gait moving pictures is proposed [25].

Image classification considering probability density function based on simplified beta distribution is proposed [26]. Also, Maximum Likelihood (MLH) classification based on classified result of boundary mixed pixels for high spatial resolution of satellite images is proposed [27]. Meanwhile, context classification based on mixing ratio estimation by means of inversion theory is proposed [28]. On the other hand, optimum spatial resolution of satellite-based optical sensors for maximizing classification performance is discussed [29]. Combined non-parametric and parametric classification method depending on normality of PDF of training samples is proposed [30].

### III. EXPERIMENT

#### A. Data used

180 of training samples and 30 of test samples are used. A small portion of the data used as a good data is shown in Fig. 2(a) while those for a bad data are shown in Fig. 2(b), respectively. There are 24 images in total, including two good and bad training data and two good and bad test data, including data augmentation (noise, skew).

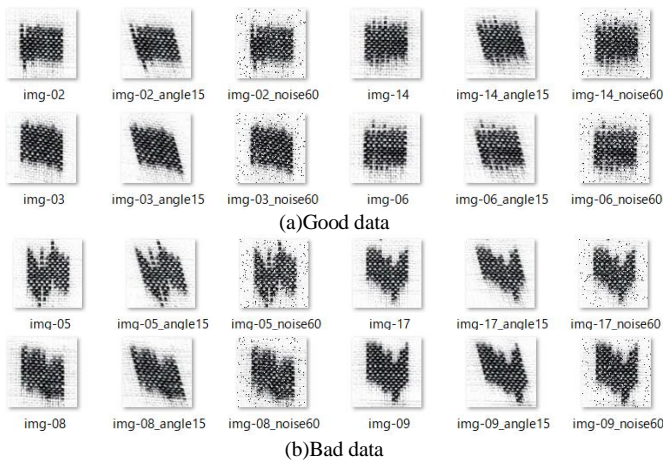


Fig. 2. A portion of good and bad data used.

#### B. Classification Methods for Comparison

There are the following methods for comparing classification performance of the good or bad categories,

- Extra Trees Classifier
- Logistic Regression
- Ridge Classifier

- Random Forest Classifier
- Light Gradient Boosting Machine
- Ada Boost Classifier
- SVM - Linear Kernel
- Gradient Boosting Classifier
- K Neighbors Classifier
- Naive Bayes
- Linear Discriminant Analysis
- Decision Tree Classifier
- Quadratic Discriminant Analysis
- Dummy Classifier

These are typical and widely used classification methods. Therefore, details of explanation of the method are not necessary.

#### C. Evaluated Performances

The following typical and widely used classification performances are evaluated for comparison,

- Accuracy
- AUC
- Recall
- Precision
- F1 score
- Kappa value
- MCC
- TT (Sec)

These are typical and widely used classification performances. Therefore, it seems that details of explanation of the performances are not necessary.

#### D. Evaluation Results

Evaluation results are summarized in Table I. Cells shaded yellow have the best performance. Almost all the classification performances of the Extra Trees Classifier show the best performance except "Recall". Total time required for learning and classification is also not so bad either. Extra Trees are similar to Random Forests. The points to construct multiple trees are the same, but one of the features (Gini coefficient, entropy) at which the node (leaf) of the tree is divided is randomly selected.

In the decision tree, when dividing the feature axis, the feature that maximizes the gain based on the feature amount (Gini coefficient, entropy), etc., and the threshold for the division are selected. Extra-Trees randomly select them. Prepare multiple random trees and bag them in the same way as Random Forest.

TABLE I. CLASSIFICATION PERFORMANCE EVALUATED

Model	Accuracy	AUC	Recall	Precision	F1	Kappa	MCC	TT (Sec)
Extra Trees Classifier	0.9176	0.9783	0.8786	0.9689	0.9122	0.8354	0.8522	0.6090
Logistic Regression	0.9110	0.9737	0.8679	0.9607	0.9078	0.8224	0.8317	1.0200
Ridge Classifier	0.9105	0.0000	0.8786	0.9532	0.9091	0.8206	0.8319	0.5470
Random Forest Classifier	0.8390	0.9417	0.8232	0.8842	0.8359	0.6783	0.6981	0.6020
Light Gradient Boosting Machine	0.8138	0.9084	0.8393	0.8253	0.8210	0.6260	0.6475	0.5670
Ada Boost Classifier	0.8133	0.8434	0.8018	0.8381	0.8105	0.6277	0.6430	0.5950
SVM - Linear Kernel	0.8067	0.0000	0.8143	0.8368	0.8045	0.6136	0.6432	0.5430
Gradient Boosting Classifier	0.7924	0.9156	0.8107	0.8040	0.7955	0.5832	0.5991	0.5700
K Neighbors Classifier	0.7910	0.9162	0.7018	0.8530	0.7510	0.5842	0.6015	0.9240
Naive Bayes	0.7438	0.7821	0.7304	0.7696	0.7460	0.4872	0.4924	0.5760
Linear Discriminant Analysis	0.7219	0.8245	0.6589	0.7731	0.6906	0.4447	0.4655	0.5730
Decision Tree Classifier	0.7157	0.7143	0.7714	0.7064	0.7271	0.4298	0.4471	0.5630
Quadratic Discriminant Analysis	0.5229	0.5205	0.5839	0.5260	0.5512	0.0406	0.0366	0.5730
Dummy Classifier	0.5133	0.5000	1.0000	0.5133	0.6783	0.0000	0.0000	0.5670

TABLE II. PARAMETERS FOR EXTRA TREES LEARNING PROCESSES

Parameters	
bootstrap	False
ccp_alpha	0.0
class_weight	{}
criterion	gini
max_depth	7
max_features	sqrt
max_leaf_nodes	None
max_samples	None
min_impurity_decrease	0.001
min_samples_leaf	5
min_samples_split	2
min_weight_fraction_leaf	0.0
n_estimators	110
n_jobs	-1
oob_score	False
random_state	123
verbose	0
warm_start	False

```
Pipeline(memory=FastMemory(location=C:\Users\Yshima\AppData\Local\Temp\joblib),
steps=[('numerical_imputer',
TransformerWrapper(exclude=None,
include=['pixel_1', 'pixel_2', 'pixel_3',
'pixel_4', 'pixel_5', 'pixel_6',
'pixel_7', 'pixel_8', 'pixel_9',
'pixel_10', 'pixel_11', 'pixel_12',
'pixel_13', 'pixel_14', 'pixel_15',
'pixel_16', 'pixel_17', 'pixel_18',
'pixel_19', 'pi...
ExtraTreesClassifier(bootstrap=False, ccp_alpha=0.0,
class_weight={}, criterion='gini',
max_depth=7, max_features='sqrt',
max_leaf_nodes=None, max_samples=None,
min_impurity_decrease=0.001,
min_samples_leaf=5, min_samples_split=2,
min_weight_fraction_leaf=0.0,
n_estimators=110, n_jobs=-1,
oob_score=False, random_state=123,
verbose=0, warm_start=False)],
verbose=False)
```

Fig. 3. Finalized parameters of the extra trees classifier.

Bagging is short for bootstrap aggregating. As the name suggests, the training data used for each learner is obtained by bootstrap sampling, and the learned learner is used for prediction and the final ensemble is performed.

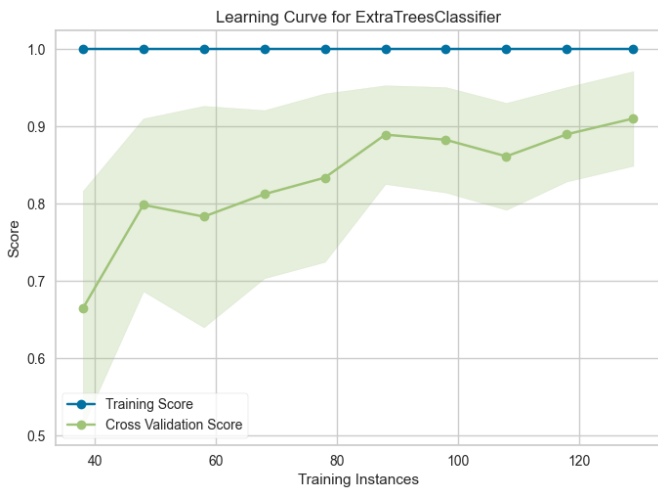
Bootstrap sampling is a method of resampling by randomly extracting some data from the population while allowing overlaps when there is data to be used as a population.

Ensemble learning in machine learning is a method of generating a single learning model by fusing multiple models (learners). A feature of Extra-Trees learning is that each tree is trained without bootstrap sampling, that is, without resampling by randomly extracting some data from the population while allowing for duplication. Use all data. Let it learn from all the data. Table II shows the parameters for Extra Trees learning processes. After the above mentioned hyperparameter tuning, the Extra Trees Classifier is finalized as shown in Fig. 3. Accuracy was specified as an optimization metric. Accuracy improved after 100 trials, but there is no guarantee that the parameters are optimal because they are optimized by random grid search.

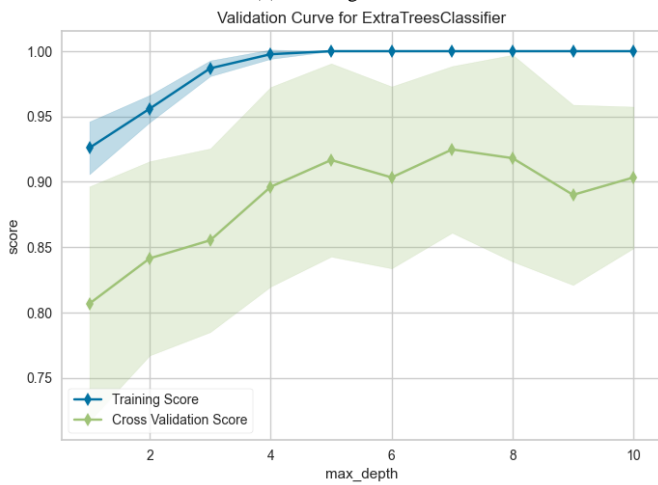
Because it is simple, processing speed is very high, and the classification accuracy is better than Random Forest, a representative machine learning library for Python. The Extra Trees Classifier when using scikit-learn looks like this: When you think of machine learning, you might think of using complex formulas or something that seems difficult, but with scikit-learn you can try out machine learning very easily.

The learning curve of the Extra Trees Classifier is shown in Fig. 4(a) while the validation curve is shown in Fig. 4(b), respectively.

N-fold cross validation with shuffle and stratification (for classification tasks). Different hyperparameters for each algorithm were checked during the training. For binary classification the Area Under ROC Curve (AUC) metric was used. Table III shows the results of the n-fold cross validation while Fig. 5 shows the AUC.



(a) Learning Curve



(b) Validation Curve

Fig. 4. Learning and validation curves of the extra trees classification.

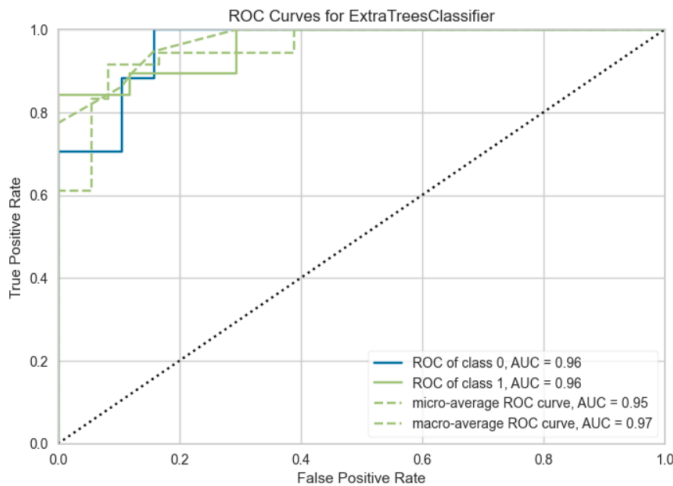


Fig. 5. Evaluated AUC.

TABLE III. THE RESULTS FROM THE N-FOLD CROSS VALIDATION

Fold	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC
0	0.9333	0.9821	1.0000	0.8889	0.9412	0.8649	0.8729
1	0.9333	0.9107	0.8750	1.0000	0.9333	0.8673	0.8750
2	0.8000	0.9464	0.6250	1.0000	0.7692	0.6087	0.6614
3	0.8667	0.9643	1.0000	0.8000	0.8889	0.7273	0.7559
4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
5	0.8571	1.0000	0.7143	1.0000	0.8333	0.7143	0.7454
6	0.9286	0.9796	0.8571	1.0000	0.9231	0.8571	0.8660
7	0.8571	1.0000	0.7143	1.0000	0.8333	0.7143	0.7454
8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Mean	0.9176	0.9783	0.8786	0.9689	0.9122	0.8354	0.8522
Std	0.0669	0.0285	0.1387	0.0653	0.0761	0.1327	0.1163

Also, confusion matrix of the Extra Trees Classifier is shown in Fig. 6. In the Fig. 6, the number of “Good” samples is 17 while the number of “Bad” samples is 19. Of the 180 training data, 20% are used as validation data, so there are 36 validation data. Of the 36 cards, 17 are 0: good and 19 are 1: bad. All good answers were correct (17 0), and three bad answers were incorrect (3 16). It is such a confusion matrix.

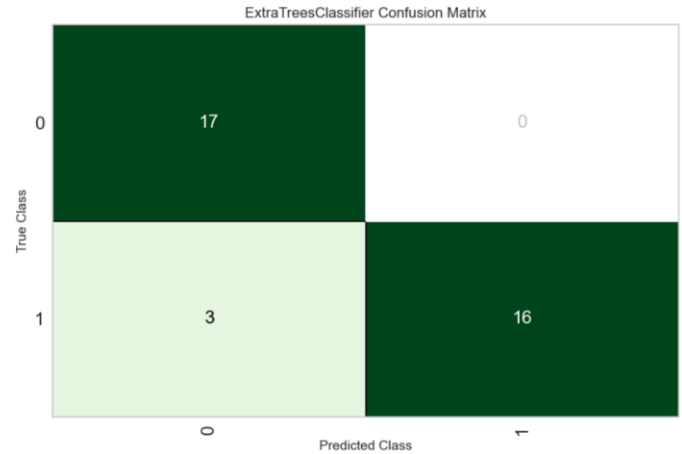


Fig. 6. Confusion matrix of the extra trees classifier.

#### IV. CONCLUSION

A method for hyperparameter tuning of image classification with PyCaret is proposed. The application example compares 14 classification methods and confirms that Extra Trees Classifier has the best performance among them, AUC=0.978, Recall=0.879, Precision=0.969, F1=0.912, Time=0.609 bottom. The Extra Trees Classifier produces a large number of decision trees, similar to the random forest algorithm, but with random sampling of each tree and no permutation.

This creates a dataset for each tree containing unique samples, and from the ensemble set of features a certain number of features are also randomly selected for each tree. The most important and unique property of the Extra Trees Classifier is that the feature split values are chosen randomly. Instead of using Gini or entropy to split the data to compute locally optimal values, the algorithm randomly selects split values.

This makes the tree diverse and uncorrelated, i.e. the diversity of each tree. Therefore, it is considered that the classification performance is better than other classification methods. Parameter tuning of Extra Trees Classifier was performed, and training performance, test performance, ROC curve, accuracy rate characteristics, etc. were evaluated.

#### FUTURE RESEARCH WORKS

Further investigations are required to identify alternative prediction methods that may lead to more accurate results.

#### ACKNOWLEDGMENT

The authors would like to thank to Professor Dr. Hiroshi Okumura and Professor Dr. Osamu Fukuda of Saga University for their valuable discussions.

#### REFERENCES

- [1] Kohei Arai, Method for Training and White Boxing DL, BDT, Random Forest and Mind Maps Based on GNN, Appl. Sci. 2023, 13, 4743. <https://doi.org/10.3390/app13084743/>, 2023.
- [2] Jin Shimazoe, Kohei Arai, Mariko Oda, Jewon Oh, Method for 1/ Fluctuation Component Extraction from Images and Its Application to Improve Kurume Kasuri Quality Estimation, International Journal of Advanced Computer Science and Applications, 13, 11, 465-471, 2022.
- [3] Kohei Arai, Classification by Re-Estimating Statistical Parameters Based on Auto-Regressive Model, Canadian Journal of Remote Sensing, Vol.16, No.3, pp.42-47, Jul.1990.
- [4] Kohei Arai, Multi-Temporal Texture Analysis in TM Classification, Canadian Journal of Remote Sensing, Vol.17, No.3, pp.263-270, Jul.1991.
- [5] Kohei Arai, Maximum Likelihood TM Classification Taking into account Pixel-to-Pixel Correlation, Journal of International GEOCATO, Vol.7, pp.33-39, Jun.1992.
- [6] Kohei Arai, A Supervised TM Classification with a Purification of Training Samples, International Journal of Remote Sensing, Vol.13, No.11, pp.2039-2049, Aug.1992.
- [7] Kohei Arai, TM Classification Using Local Spectral Variability, Journal of International GEOCATO, Vol.7, No.4, pp.1-9, Oct.1992.
- [8] Kohei Arai, A Classification Method with Spatial Spectral Variability, International Journal of Remote Sensing, Vol.13, No.12, pp.699-709, Oct.1992.
- [9] M.Moriyama and Kohei Arai, An Inversion for Emissivity-Temperature Separation with ASTER Data, Advances in Space Research, Vol.14, No.3, pp.67-70, Jul.1993.
- [10] Kohei Arai, TM Classification Using Local Spectral Variability, International Journal of Remote Sensing, Vol.14, No.4, pp.699-709, 1993.
- [11] Kohei Arai, Application of Inversion Theory for Image Analysis and Classification, Advances in Space Research, Vol.21, 3, 429-432, 1998.
- [12] Kohei Arai and J.Wang, Polarimetric SAR image classification with maximum curvature of the trajectory in eigen space domain on the polarization signature, Advances in Space Research, 39, 1, 149-154, 2007.
- [13] Hiroshi Okumura, Makoto Yamaura and Kohei Arai, A hybrid supervised classification method for multi-dimensional images using color and textural features, Journal of the Institute of Image Electronics Engineers of Japan, 38, 6, 872-882, 2009.
- [14] Kohei Arai, Rosa Andrie Asmara, Human gait gender classification using 3D discrete wavelet transformation feature extraction, International Journal of Advanced Research in Artificial Intelligence, 3, 2, 12-17, 2014.
- [15] Kohei Arai, Polarimetric SAR image classification with high frequency component derived from wavelet multi resolution analysis: MRA, International Journal of Advanced Computer Science and Applications, 2, 9, 37-42, 2011.
- [16] Kohei Arai Comparative study of polarimetric SAR classification methods including proposed method with maximum curvature of trajectory of backscattering cross section in ellipticity and orientation angle space, International Journal of Research and Reviews on Computer Science, 2, 4, 1005-1009, 2011.
- [17] Kohei Arai, Rosa Andrie, Human gait gender classification using 2D discrete wavelet transforms energy, International Journal of Computer Science and Network Security, 11, 12, 62-68, 2011.
- [18] Kohei Arai, R.A.Asunara, Human gait gender classification in spatial and temporal reasoning, International Journal of Advanced Research in Artificial Intelligence, 1, 6, 1-6, 2012.
- [19] Kohei Arai, Comparative study on discrimination methods for identifying dangerous red tide species based on wavelet utilized classification methods, International Journal of Advanced Computer Science and Applications, 4, 1, 95-102, 2013.
- [20] Kohei Arai, Multi spectral image classification method with selection of independent spectral features through correlation analysis, International Journal of Advanced Research in Artificial Intelligence, 2, 8, 21-27, 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, Rosa Andrie Asmara, Gender classification method based on gait energy motion derived from silhouettes through wavelet analysis of human gait moving pictures, International Journal of Information Technology and Computer Science, 6, 3, 1-11, 2014.
- [23] Kohei Arai, Rosa Andrie Asmara, Human gait skeleton model acquired with single side video camera and its application and implementation for gender classification, Journal of the Image Electronics and Engineering Society of Japan, Transaction of Image Electronics and Visual Computing, 1, 1, 78-87, 2013.
- [24] Kohei Arai, Rosa Andrie Asmara, Human gait skeleton model acquired with single side video camera and its application and implementation for gender classification, Journal of the Image Electronics and Engineering Society of Japan, Transaction of Image Electronics and Visual Computing, 1, 1, 78-87, 2014.
- [25] Kohei Arai, Rosa Andrie Asmara, Gender classification method based on gait energy motion derived from silhouette through wavelet analysis of human gait moving pictures, International Journal of Information technology and Computer Science, 5, 5, 12-17, 2013.
- [26] Kohei Arai, Image classification considering probability density function based on Simplified beta distribution, International Journal of Advanced Computer Science and Applications IJACSA, 11, 4, 481-486, 2020.
- [27] Kohei Arai, Maximum Likelihood Classification based on Classified Result of Boundary Mixed Pixels for High Spatial Resolution of Satellite Images, International Journal of Advanced Computer Science and Applications, Vol. 11, No. 9, 24-30, 2020.
- [28] Kohei Arai, Context Classification based on Mixing Ratio Estimation by Means of Inversion Theory, International Journal of Advanced Computer Science and Applications, Vol. 11, No. 12, 44-50, 2020.
- [29] Kohei Arai, Optimum Spatial Resolution of Satellite-based Optical Sensors for Maximizing Classification Performance, Journal of Advanced Computer Science and Applications, Vol. 12, No. 2, 363-369, 2021.

- [30] Kohei Arai, Combined Non-parametric and Parametric Classification Method Depending on Normality of PDF of Training Samples, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 12, No. 5, 310-316, 2021.

#### AUTHORS' 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>

Jin Shimazoe, He received BE degree in 2022. He also received the IEICE Kyushu Section Excellence Award. He is currently working on research that uses image processing and image recognition in Master's Program at Kurume Institute of Technology.

Mariko Oda, She graduated from the Faculty of Engineering, Saga University in 1992, and completed her master's and doctoral studies at the Graduate School of Engineering, Saga University in 1994 and 2012, respectively. She received Ph.D (Engineering) from Saga University in 2012. She also received the IPSJ Kyushu Section Newcomer Incentive Award. In 1994, she became an assistant professor at the department of engineering in Kurume Institute of Technology; in 2001, a lecturer; from 2012 to 2014, an associate professor at the same institute; from 2014, an associate professor at Hagoromo university of International studies; from 2017 to 2020, a professor at the Department of Media studies, Hagoromo university of International studies. In 2020, she was appointed Deputy Director and Professor of the Applied of AI Research Institute at Kurume Institute of Technology. She has been in this position up to the present. She is currently working on applied AI research in the fields of education.



# A Novel Artifact Removal Strategy and Spatial Attention-based Multiscale CNN for MI Recognition

Duan Li, Peisen Liu, Yongquan Xia

School of Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou, Henan, China

**Abstract**—The brain-computer interface (BCI) based on motor imagery (MI) is a promising technology aimed at assisting individuals with motor impairments in regaining their motor abilities by capturing brain signals during specific tasks. However, non-invasive electroencephalogram (EEG) signals collected using EEG caps often contain large numbers of artifacts. Automatically and effectively removing these artifacts while preserving task-related brain components is a key issue for MI de-coding. Additionally, multi-channel EEG signals encompass temporal, frequency and spatial domain features. Although deep learning has achieved better results in extracting features and de-coding motor imagery EEG (MI-EEG) signals, obtaining a high-performance network on MI that achieves optimal matching of feature extraction, thus classification algorithms is still a challenging issue. In this study, we propose a scheme that combines a novel automatic artifact removal strategy with a spatial attention-based multiscale CNN (SA-MSCNN). This work obtained independent component analysis (ICA) weights from the first subject in the dataset and used K-means clustering to determine the best feature combination, which was then applied to other subjects for artifact removal. Additionally, this work designed an SA-MSCNN which includes multiscale convolution modules capable of extracting information from multiple frequency bands, spatial attention modules weighting spatial information, and separable convolution modules reducing feature information. This work validated the performance of the proposed model using a real-world public dataset, the BCI competition IV dataset 2a. The average accuracy of the method was 79.83%. This work conducted ablation experiments to demonstrate the effectiveness of the proposed artifact removal method and SA-MSCNN network and compared the results with outstanding models and state-of-the-art (SOTA) studies. The results confirm the effectiveness of the proposed method and provide a theoretical and experimental foundation for the development of new MI-BCI systems, which is very useful in helping people with disabilities regain their independence and improve their quality of life.

**Keywords**—Motor Imagery (MI); Brain Computer Interface (BCI); EEG signal; artifact removal; spatial attention; Convolutional Neural Network (CNN)

## I. INTRODUCTION

Brain-computer interfaces (BCIs) have the capability of translating brain activity signals into commands to control external devices or communicate with the external environment [1]. One of the most common paradigms of BCIs is motor imagery (MI), which involves mentally simulating motor commands of specific body parts. The generation of MI signals is possible even in the presence of disabilities as the

corresponding brain region is functioning properly. MI-BCIs have shown promising results in areas such as communication, control and rehabilitation [2-7]. Electroencephalography (EEG) is widely used for MI-BCI systems due to its convenience and low physical and psychological stress on the subject [8]. However, EEG signals can be affected by environmental factors, which can generate artifacts. The background noise will distort the signal of interest and consequently reduce the MI recognition accuracy. In recent years, deep learning has gradually received attention for MI recognition. However, extracting appropriate spatial and temporal information from EEG signals has always been a significant challenge, whether in deep learning or traditional studies.

Due to the fact that EEG signals are collected by electrodes in contact with the scalp, the signal-to-noise ratio of the EEG is relatively low. During the acquisition process, the EEG signals are easily contaminated by various factors, resulting in artifacts in the signal, such as eye movements, muscle movements and electric line noise [9,10]. Therefore, some studies exploring different types of artifact features and removal methods have achieved certain results. Independent component analysis (ICA) [11] is a widely used method for artifact removal in MI recognition. ICA can separate a specified number of components from the input signal. Typically, experienced researchers identify and exclude the components that are considered artifacts and then the remaining components are used to reconstruct the signal for further analysis. This approach has achieved some success in various studies [12-14]. However, classifying components extracted by ICA will require much time and effort from professionals, which is not feasible for large datasets. Therefore, the automatic classification of ICA components has been proposed in some literature. For example, Hesam et al. [15] used three features-based K-means clustering methods to automatically cluster and differentiate artifacts from brain states in ICA components, achieving a 3.95% accuracy improvement on the Physionet dataset. However, there is still a lack of effective exploration of different feature methods and the impact of different features on clustering is not yet clear. Therefore, further research in this area is necessary.

Traditional convolutional neural networks extract and learn features through convolutions. In contrast to the two-dimensional convolutions commonly used in image applications, one-dimensional convolutions are often employed in MI-related research to capture temporal information and electrode information from EEG signals. For example, well-known models such as EEGNet [16] and DeepConvNet [17], utilize one-dimensional convolutions to extract

information from the signals. Recently, the TS-SEFFNet [18] also incorporates one-dimensional convolutions with multiple scales, which has been proven to be effective. The multi-scale module can effectively concentrate on multiple dimensional information. How to improve the useful multi-dimensional features and suppress useless information is a key issue for MI recognition. Attention mechanism [19], after its proposal in 2014, which has been widely used in both the computer vision (CV) [20] and natural language processing (NLP) [21] fields, provides an efficient way for multiple channel EEG recognition. The Attention mechanism mainly focuses limited attention on important information, thereby saving resources and quickly obtaining the most relevant information. There are three main types of attention models for multi-channel EEG: spatial attention, channel attention and a combination of spatial and channel attention [22]. Spatial attention pays attention to specific regions within the spatial domain, allowing the model to prioritize important spatial information. Channel attention assigns different weights to different channels based on their attention values, facilitating the model to focus on important channel-wise features. However, in MI-BCI research, constructing a robust CNN is still a challenging problem. The utilization of various modules in the network structure still requires exploration.

In this paper, a novel features selection ICA and K-means-based automatic artifact removal method and an end-to-end CNN that includes multi-scale convolutions and spatial attention mechanism were proposed to overcome the problem of low classification accuracy and enhance the network's robustness. It is well known that the artifacts can lead to the acquisition of task-irrelevant features by subsequent classifiers, resulting in a decrease in classification accuracy. Therefore, this work proposes an automatic artifact removal method that combines ICA and clustering algorithms. For the subsequent network architecture, this work introduces a modularized network called SA-MSCNN. It utilizes multi-scale block and spatial attention modules to extract different types of information. The method takes into account the deep network's ability to learn different features from the data and offers a novel approach for classifying motor imagery. This work evaluates the performance of the proposed method using 5184 trials with 22 EEG channels from nine subjects in the BCI-IV-2a dataset.

The remainder of this paper is organized as follows. Section II introduces the method for this study which includes the proposed artifact removal method and SA-MSCNN. In Section III, this work introduces the dataset and the evaluation metrics used in this work. And also describes preprocessing steps, experimental parameter settings and the experimental results. Section IV is the discussion and Section V concludes the paper.

## II. METHOD

This section describes the method this work proposed in this paper. First, this work gives an overall framework of the proposed method and a brief introduction to the workflow. Then, this work describes in detail the proposed ICA+K-means artifact removal method and the SA-MSCNN. Finally, this work shows the training strategy for the proposed network.

### A. Overall Framework

The proposed method is shown in Fig. 1. This work first pre-processed the data for all subjects and then performed artifact removal using the proposed ICA+K-means method. For each subject, this work pre-trained the proposed Spatial Attention-based Multi Scale Convolution Neural Network (SA-MSCNN) using data from all the subjects except the specific subject, then saved the SA-MSCNN's weights. Finally, the SA-MSCNN model was trained and adjusted by the specific subject data (on the right side of Fig. 1) and the following experiments were performed to obtain the classification results for comparison.

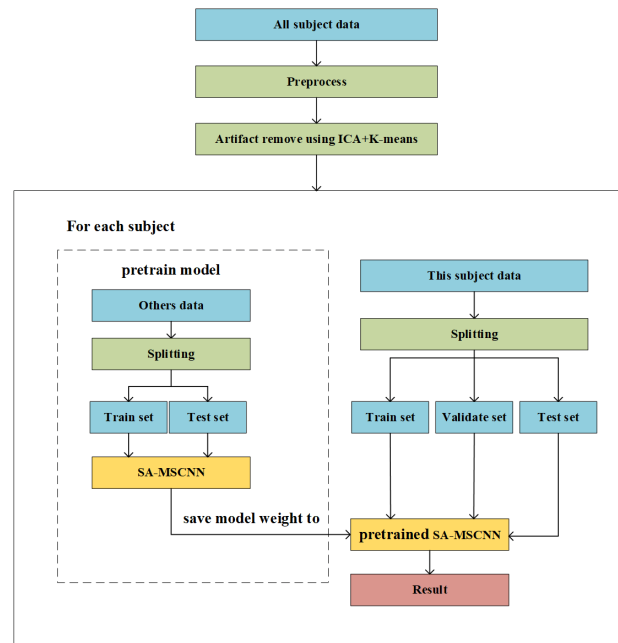


Fig. 1. Block diagram of the proposed framework.

### B. Artifact Removal using ICA+K-means

Independent Component Analysis (ICA) is a statistical method used for signal processing and data analysis. It is widely used in MI to remove artifacts from EEG signals [23]. Unlike other traditional signal analysis methods such as Principal Component Analysis (PCA), ICA emphasizes the independence rather than the correlation of the signals. ICA decomposes signals according to Eq. (1) and (2).

$$\mathbf{X}_{m \times n} = \mathbf{A}_{m \times m} \mathbf{Y}_{m \times n} \quad (1)$$

$$\mathbf{Y}_{m \times n} = \mathbf{W}_{m \times m} \mathbf{X}_{m \times n} \quad (2)$$

Where X is the input EEG signal, A represents the mixing matrix, Y represents the original signal sources and W represents the inverse matrix of A. Here, m and n represent the number of EEG channels and the number of samples, respectively. The objective of the ICA algorithm is to find the weight matrix W that will decompose the EEG signals into ICs assuming temporal and spatial independence. Most artifact ICs in MI EEG signals are caused by eye movements, muscle movements and electric line noise [9, 10]. Traditionally, manual observation of ICs are used to differentiate artifact

components from brain-related components and then remove the artifact components. This approach is feasible for datasets with a small number of subjects or a small amount of data for each subject. However, with the recent development of EEG measurement devices and experimental paradigms, large datasets like OpenBMI have emerged [24]. As the amount of data in the dataset increases, the manual selection of ICs becomes cumbersome and time-consuming. Therefore, the development of automated methods for ICs selection is necessary.

To address the above-mentioned issues, Hesam et al. [15] proposed using the K-means clustering method for component selection on extracted ICs from datasets including Physionet. They extracted three specific features from ICs and performed clustering, designating the class with the highest variance as the artifact class, and removing the components in that class. However, they only considered three specific feature selections and did not demonstrate the clustering results. Moreover, they did not explore the impact of different feature selections and multi-class datasets on K-means clustering from a broader perspective. Therefore, this study focuses on studying the impact of different feature selections on K-means clustering using representative datasets such as BCI-IV-2a. It provides recommendations for feature combinations and presents brain maps and ablation experiments after clustering. The seven selected statistical measures are variance, covariance, inverse covariance, correlation coefficient, kurtosis, skewness, and quartile range. Since the extracted ICs are vectors of certain lengths, these seven statistical measures can extract corresponding features of the ICs, as shown in Eq. (3) to (9) for their extraction method.

$$\sigma^2 = \frac{\sum (Y_i - \mu)^2}{n} \quad (3)$$

$$Cov(Y_i, Y_j) = E(Y_i, Y_j) - E(Y_i)E(Y_j) \quad (4)$$

$$invCov(Y_i, Y_j) = Cov(Y_i, Y_j)^{-1} \quad (5)$$

$$p = \frac{Cov(Y_i, Y_j)}{\sigma_{Y_i} \sigma_{Y_j}} \quad (6)$$

$$Kurt = \frac{\frac{1}{n} \sum_{i=1}^n (Y_i - \mu)^4}{\sigma^4} \quad (7)$$

$$Skew = \frac{\frac{1}{n} \sum_{i=1}^n (Y_i - \mu)^3}{\sigma^3} \quad (8)$$

$$IQR = Y_{75\%} - Y_{25\%} \quad (9)$$

Where  $\sigma^2$  represents variance,  $Cov$  represents covariance,  $E(A, B)$  represents the expected value of a function involving two random variables,  $A$  and  $B$ , with their respective probability distributions,  $invCov$  represents inverse covariance,  $p$  represents correlation coefficient,  $Kurt$  represents kurtosis,  $Skew$  represents skewness, and  $IQR$  represents interquartile range.  $Y_i$  represents the  $i$ -th sample,  $\mu$  represents the sample mean,  $n$  represents the total number of samples, and  $Y_{75\%}/Y_{25\%}$  represents the values at the 75th/25th percentile when the data is sorted in ascending order.

K-means is an unsupervised clustering algorithm where the parameter "K" in its name is set by the user to determine the number of clusters in the final result. The K-means algorithm iteratively maximizes the similarity among data points within each cluster while minimizing the similarity between clusters until the cluster centers no longer change or the predetermined number of iterations is reached. The most commonly used similarity measure is distance, such as Euclidean distance or Manhattan distance. The choice of K significantly impacts the clustering result and therefore, K-means clustering may converge to a locally optimal solution.

The different feature combinations for K-means algorithm will result in different clustering results. How to select optimal features for the automation of artifacts ICs identification is a key issue. Therefore, this study explores the different combinations of the aforementioned seven features and then adaptively achieves the best clustering results for denoising. The process is illustrated in Fig. 2. First, we extract the original ICA weights from the raw EEG signals of the first subject in the dataset.

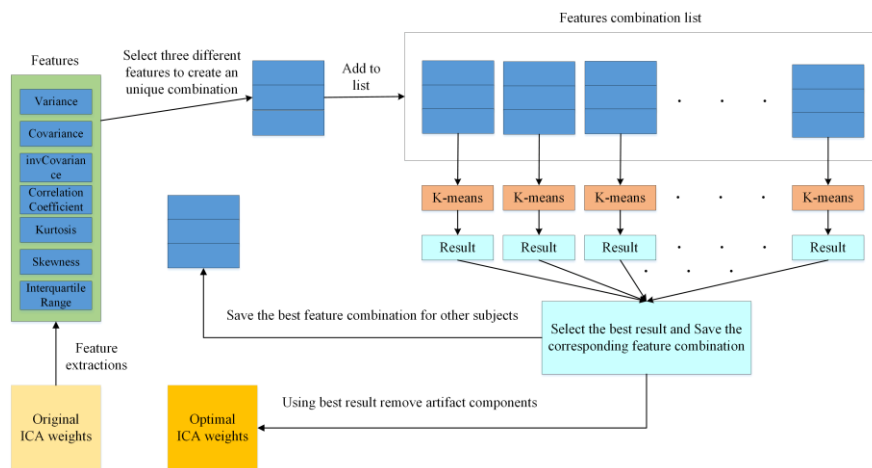


Fig. 2. Proposed ICA+K-means artifacts removal framework.

These weights are obtained by labeling the brain components and artifact components by experienced experts. Next, seven features were extracted from ICA weights. For these features, this work selects three different features to form a unique combination and then adds them to a feature's combination list. Each combination in the list is distinct from the others, resulting in a total of 35 unique feature combinations. For each combination, the K-means algorithm is used to cluster using the three combined features. Then evaluate the clustering results based on the previous labels. Finally, select the best feature combination that achieves the optimal results. This work uses the best feature combination to obtain the optimal ICA weights in the following experiments for removing artifacts from the rest subjects. The proposed method is capable of automatically removing artifacts, reducing the expenditure of time and effort.

### C. Spatial Attention-based Multi Scale Convolution Neural Network (SA-MSCNN)

The proposed SA-MSCNN is an end-to-end CNN, composed primarily of a multi scale temporal convolution block, a spatial convolution block, a spatial attention block, a separable convolution block and a classification block. The network structure is illustrated in Fig. 3.

As the 22 leads EEG data sampled at 250Hz and each sample has a length of 2.5s, the size of the sample is  $22 \times 625$ . The multi scale temporal convolution blocks and spatial convolution blocks of SA-MSCNN primarily capture the temporal and spatial features of the input EEG signals. The filters for the multi-scale temporal convolution block and spatial convolution block are set as  $[1, k]$  and  $[c, 1]$  respectively, where  $k$  and  $c$  are the lengths of the filters. Then, 2D convolution is performed accordingly. As mentioned in FBCSP [25], there may be some fluctuations in the MI frequency bands for different subjects. Therefore, to improve the classification accuracy, this work incorporates multi-scale convolutions with filters of different lengths in the temporal dimension to obtain information at different time scales. The obtained multi-scale features are then concatenated, normalized and passed through the spatial convolution blocks to extract spatial-scale information.

The subsequent Spatial Attention Block primarily utilizes the spatial attention module to obtain a refined feature map, further enhancing the model's attention and perception abilities in the spatial domain. The input feature, after undergoing max pooling and average pooling, has a shape of  $[\text{batch\_size}, \text{input\_channels}, \text{height}, \text{width}]$ . Then, a convolution operation is performed, resulting in a feature map with a shape of  $[\text{batch\_size}, 1, \text{height}, \text{width}]$ . The values of the feature map are then mapped to a range between 0 and 1 using the Sigmoid function, achieving attention weights. Finally, these weights are multiplied element-wise with the original feature map, resulting in a weighted feature map, which is referred to as the refined feature map. In traditional feature extraction networks, the features at each position in the feature map are treated equally, without considering the importance of different positions. However, in real-world scenarios, the contribution of information from different positions varies. The spatial attention module introduces different weights to each position in the feature map, allowing the network to focus more on important positions and regions. This enables the network to capture richer and more accurate feature information thus improving the overall performance. Moreover, since the spatial attention module is introduced in the middle module, it does not directly process the raw EEG signals. It only requires the intermediate weights for feature extraction, thereby almost not increasing the complexity and computational cost of the network.

The refined feature map is performed batch normalization and ELU activation. Average pooling is applied to reduce the size of the feature map, while dropout is employed to prevent overfitting of the model. Then the feature map is fed into the depthwise separable convolution block. The depthwise separable convolution consists of depthwise convolution and pointwise convolution, with filters set as  $[1, k]$  and  $[1, 1]$ , respectively. These convolutions aim to further shrink the feature map. Subsequently, batch normalization, ELU activation, average pooling and dropout are performed and the results are passed to the classification block. In the classification block of SA-MSCNN, the input features are flattened and then fed into a fully connected layer to compute the final output classification label.

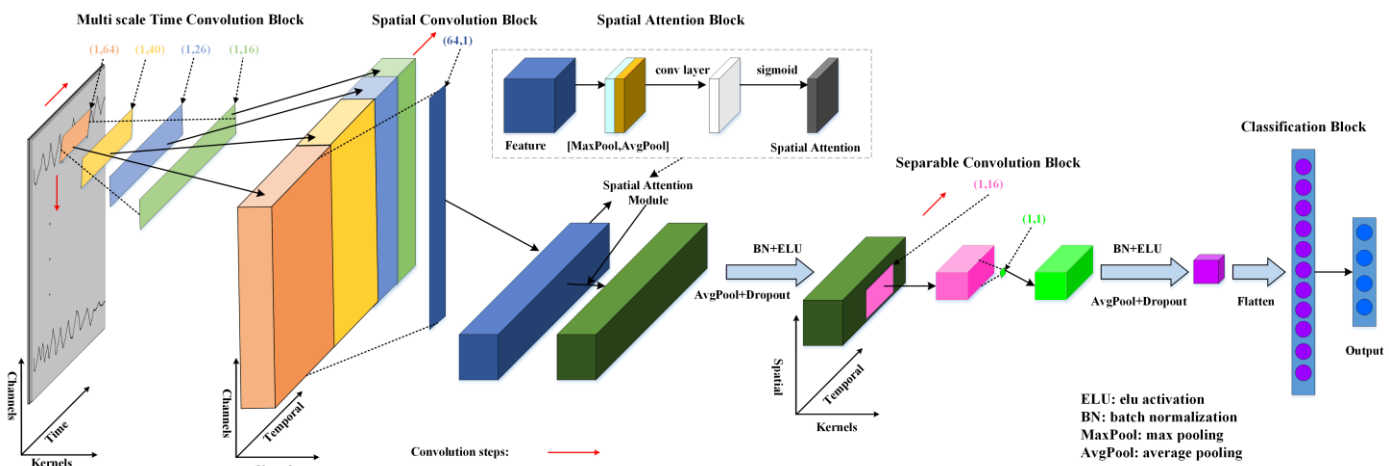


Fig. 3. Proposed SA-MSCNN architecture.

#### D. Training Strategy

For each subject, SA-MSCNN is pre-trained using data from all other subjects except the specific one and then trained, validated, and tested on the pre-trained model using the data of this specific subject. The reason for using the pre-training step is inspired by the idea of transfer learning. This training strategy aims to enable the model to learn features from the entire dataset as much as possible and increase the amount of data used for training. This allows the model parameters to converge faster when the data of the specific subject is fed into the pre-trained SA-MSCNN.

### III. EXPERIMENTS AND RESULTS

#### A. Data Sources and Evaluation Metrics

In this experiment, dataset 2a of the BCI competition IV, which contains 22 EEG channels and three monopolar EOG channels from nine subjects was used [26]. The 22 EEG electrodes were made by Ag/AgCl (with inter-electrode distances of 3.5 cm). The sampling frequency is 250Hz and bandpass filtering between 0.5Hz and 100Hz was applied when the dataset was recorded. The subjects were asked to perform four types of motor imagery tasks: left hand, right hand, tongue and both feet. Each category of the task was performed 72 times, resulting in 288 trials per session and each subject had two sessions. So, there are a total of 5184 trials in the dataset.

In this study, the performance of the proposed method was evaluated using its accuracy (represented by the symbol  $Acc$ ) and Kappa value (represented by the symbol  $Kappa$ ), defined by Eq. (10) and (11).

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

$$Kappa = 1 - \frac{1 - P_o}{1 - P_e} = \frac{P_o - P_e}{1 - P_e} \quad (11)$$

Among them, TP, FP, TN and FN are true positive, false positive, true negative and false negative respectively.  $P_o$  represents the total classification accuracy, and  $P_e$  is the sum of the product of the ground truth and predicted numbers for each category divided by the square of the total number of samples [27].

#### B. Pre-processing

In the data preprocessing stage, this work removed three EOG channels and retained 22 EEG channels to reduce their influence. MI tasks produce event-related synchronization (ERD) and event-related desynchronization (ERS) in EEG signals, corresponding to the sensorimotor rhythms of mu (8-13Hz) and beta (18-30Hz). The ERD/ERS patterns exhibit variability across subjects. Therefore, a wide-range band-pass filter of 4-40Hz was used to minimize band-pass filtering's influence on data while retaining MI-related features. For each task, the EEG data is segmented between 0 s to 2.5s after its start. Thus, each sample size is  $22 \times 625$ . No other operations are performed because this work wants to retain as much useful information as possible.

#### C. Experimental Setup

In the artifact removal stage, for the first subject, after ICA, the first two ICs were removed, and the remaining 20 components were manually classified into artifact and non-artifact categories. Then, the proposed method was used for clustering with three clusters, as the dataset consisted of four classes. The clustering results were recorded, excluding the features where artifacts and non-artifacts were clustered into the same category. The percentage of artifact ICs in the category with the most artifact ICs and the percentage of non-artifact ICs in the category with the most non-artifact ICs were recorded separately. The two percentages were added and divided by two to obtain the final clustering distinction percentage. A higher clustering distinction percentage suggests that the corresponding feature extraction method is more effective in artifact removal. For the remaining subjects, the most effective feature extraction method was applied to remove artifacts. To compare, two experiments were also conducted on all subjects: using ICA+manual artifact removal, and without using ICA for artifact removal.

ICA can be performed using the MNE package in Python [28] or the EEGLAB toolbox in MATLAB [29]. In the ICA artifact removal stage, for the first subject's ICs, the first two ICs were removed and the remaining 20 components were manually labeled as artifact and non-artifact categories. Hesam et al. [15] only used two classes of data from the Physionet dataset for their experiment and set the clustering to two classes. In this experiment, this work used a four-class dataset. Since each session for clustering involved four classes, in order to acquire more accurate clustering results, a three-class clustering approach was opted for. Then this work employed the proposed method for all the subjects using K-means clustering with three clusters. The clustering results were recorded, excluding the feature combinations where most artifacts and most non-artifacts ICs were clustered into the same category. The percentage of artifact ICs in the category with the maximum number of artifact ICs and the percentage of non-artifact ICs in the category with the maximum number of non-artifact ICs were calculated and recorded for each feature combination. The two percentages were added and divided by two to obtain the final clustering distinction. A higher clustering distinction suggests that the corresponding feature extraction method is more effective in artifact removal. For the remaining subjects, the three most effective feature extraction methods were applied to remove artifacts. To compare, the other two experiments were also conducted on all subjects: using ICA+manual artifact removal, and without artifact removal.

In SA-MSCNN, the kernel sizes for the multi-scale time convolutional blocks are set as [1,64], [1,40], [1,26], [1,16] and the kernel size for the spatial convolutional block is set as [64,1]. The depthwise separable convolution block consists of a depthwise convolution with a kernel size of [1,16] and a pointwise convolution with a kernel size of [1,1]. Each scale in the multi-scale temporal convolutional block has eight convolutional kernels, the spatial convolutional block has 16 convolutional kernels and the depthwise separable convolution block has 16 convolutional kernels. The dropout rate is set to

0.5 and the cross-entropy loss function is used. The Adam optimizer with a learning rate of 0.001 is utilized.

During the pretraining phase, for each subject, this work combined the data of the other eight subjects, totaling 4608 trials. This combined dataset was then divided into 75% for training and 25% for testing to pre-train the model parameters of SA-MSCNN. Then, the 576 trials from the current subject were divided into 50% for training, 25% for validation and 25% for testing, which were fed into the pre-trained SA-MSCNN to obtain results. Both the pre-training and training epochs are set to 200.

All the experiments were implemented on Windows 11 with an Nvidia RTX 3060 12GB GPU and the neural network was performed on the PyTorch platform.

**D. Compared Methods**

For comparison, this work also used three recent years' outstanding open-sourced models for EEG recognition including EEGNet, DeepConvNet and TS-SEFFNet.

- EEGNet [16]: EEGNet is a compact convolutional neural network specifically designed for processing EEG data. It extracts spatial and temporal features of EEG signals through one-dimensional convolutional layers and depthwise separable convolutional layers.
- DeepConvNet [17]: DeepConvNet is a model based on a deep convolutional neural network architecture used for classifying EEG signals. It employs multiple convolutional layers, pooling layers and fully connected layers to extract high-level features.
- TS-SEFFNet [18]: TS-SEFFNet is a time-frequency-based compressed and excitatory

feature fusion network used for decoding motor imagery EEG. The network utilizes a novel time-frequency compression and excitatory feature fusion method for motor imagery EEG decoding

**E. Result of Clustering**

In Table I, Feature Combination represents the selected combinations of features. NBS MaxCate corresponds to the category where most of the artifact components are clustered. NBS MaxCatePercent represents the percentage of artifact components in this category out of the total artifact components. BS MaxCate corresponds to the category where most of the brain state components are clustered. BS MaxCatePercent represents the percentage of brain state components in this category out of the total brain state components. Clustering Distinction corresponds to the clustering distinctiveness, which is equal to half the sum of NBS MaxCatePercent and BS MaxCatePercent. In Table I, kurt represents kurtosis, skew represents skewness, cov represents covariance, iqr represents interquartile range, var represents variance, inv\_cov represents inverse covariance and corr represents correlation coefficient. Three different feature extraction methods were selected from seven available methods to form a unique combination, resulting in a total of 35 combinations. Firstly, features that cluster artifact components and brain state components into the same category were excluded. Then, features with a clustering distinctiveness of less than 60% were excluded and the results are summarized in Table I. From Table I, it can be seen that the combination of kurt, skew and cov has the highest clustering distinction. Therefore, these three feature extraction methods were used in subsequent ICA+K-means automatic clustering to remove artifact components in other subjects.

TABLE I. CLUSTER RESULT OF SUBJECT 1

Feature Combination	NBS MaxCate	NBS MaxCatePercent	BS MaxCate	BS MaxCatePercent	Clustering Distinction
kurt, skew, cov	2	80%	0	86.67%	83.34%
kurt, skew, iqr	1	60%	0	86.67%	73.34%
kurt, skew, var	1	60%	0	86.67%	73.34%
inv_cov, iqr, var	0	100%	1	46.67%	73.34%
corr, kurt, cov	0	60%	1	73.33%	66.67%
corr, kurt, var	1	60%	0	73.33%	66.67%
kurt, cov, iqr	2	40%	0	86.67%	63.34%
corr, skew, inv_cov	2	80%	0	40%	60%

TABLE II. COMPARISON OF DIFFERENT EXPERIMENTS

Method	Subjects									
	A01	A02	A03	A04	A05	A06	A07	A08	A09	Avg
base model without artifacts remove	85.71%	64.86%	93.22%	69.86%	73.71%	61.33%	88.11%	81.31%	80.47%	77.62%
base model+ICA+ Manual select	89.71%	67.25%	95.9%	75.11%	77.21%	62.79%	89.25%	82.51%	83.22%	80.33%
base model+ ICA+K-means	89.71%	64.88%	94.97%	73.19%	78.68%	60.85%	90.93%	85.31%	79.97%	79.83%

F. Ablation Experiment of Artifact Removal Method

In this section, ablation experiments without artifacts, manual select and ICA combined K-means are performed respectively. The experimental results are summarized in Table II, the second column is the experimental results without artifact removal. The "ICA+Manual select" in the table represents the experimental results where manual observation was used to remove artifacts for all subjects. In the "ICA+K-means" experiments, the artifacts removal and optimal feature combinations for the remaining subjects were according to the first subject. The results are shown in Table II and Fig. 4.

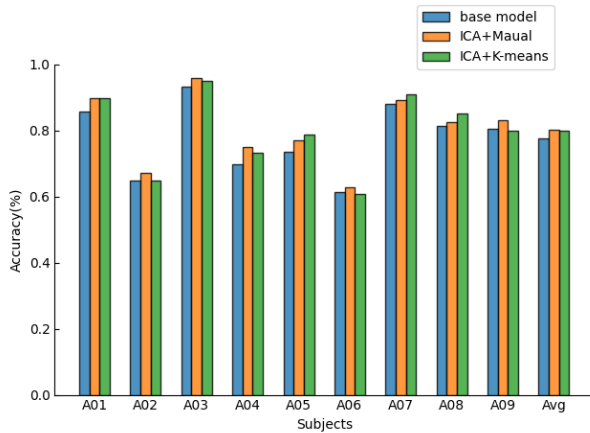


Fig. 4. Comparison of different experiments.

In the ICA+K-means method, the first subject still required manually classifying and removing artifact components. Therefore, the experimental results for the first subject, both in terms of manual artifact removal and cluster-based artifact removal, are the same. The results for the remaining subjects are different. From Fig. 4, it can be observed that both experiments using artifact removal methods outperformed the base model, which did not include artifact removal. This result demonstrates the effectiveness of artifact removal methods. Furthermore, Table II shows that the average classification accuracy for manual identification and exclusion of artifacts is 80.33%, while the average classification accuracy for using K-means clustering to identify and exclude artifacts is 79.83%, a difference of only about 1%.

Fig. 5 illustrate the training process for the third subject using the proposed method. Train\_acc and Val\_acc represent

training and validating accuracy respectively, while Train\_loss and Val\_loss represent training and validation loss. Using the pre-training strategy, the model converges speed at a fast rate.



Fig. 5. Training process of subject 3.

G. Comparative Experiment with other Networks

EEGNet, DeepConvNet, and TS-SEFFNet are popular and excellent networks in the field of BCI decoding. From Table III, DeepConvNet achieves an average accuracy of 71.99%, EEGNet achieves 72.44% and TS-SEFFNet achieves 74.71%. By using the proposed artifact removal method and SA-MSCNN, the accuracy improves to 79.83%. Furthermore, two ablation experiments were also performed: 1) Removal of the multiscale block and the spatial attention block and 2) Removal of the artifact removal module. The results demonstrate the effectiveness of the proposed method. Fig. 6 is the confusion matrix of the four methods. The values of the matrix have been normalized by rows.

As can be seen in Table III, the proposed method outperforms these comparative methods in terms of both average classification accuracy and Kappa value. This work also conducted ablation experiments to compare and prove the effectiveness of the method, both SSA-MSCNN with multi-scale convolutional block and spatial attention block and ICA+K-means artifact removal method can improve the performance of the model. In Fig. 6, the classification accuracy of the method is improved on all four classes, especially on the left and right hands.

TABLE III. COMPARISON OF OTHER METHODS

Method	Subjects' Acc										Kappa
	A01	A02	A03	A04	A05	A06	A07	A08	A09	Avg	
DeepConvNet	80.90%	52.08%	84.72%	71.18%	70.49%	55.56%	69.10%	81.94%	81.94%	71.99%	0.627
EEGNet	81.25%	50.69%	91.67%	63.89%	70.14%	59.03%	79.17%	77.98%	78.18%	72.44%	0.632
TS-SEFFNet	82.29%	49.79%	87.57%	71.74%	70.83%	63.75%	82.92%	81.53%	81.94%	74.71%	0.663
SA-MSCNN without Multiscale Block and Spatial Attention Block	83.23%	61.07%	82.53%	65.23%	70.28%	57.47%	84.90%	78.16%	75.73%	73.18%	0.642
SA-MSCNN without ICA+K-means	85.71%	64.86%	93.22%	69.86%	73.71%	61.33%	88.11%	81.31%	80.47%	77.62%	0.702
SA-MSCNN with ICA+K-means	89.71%	64.88%	94.97%	73.19%	78.68%	60.85%	90.93%	85.31%	79.97%	79.83%	0.731

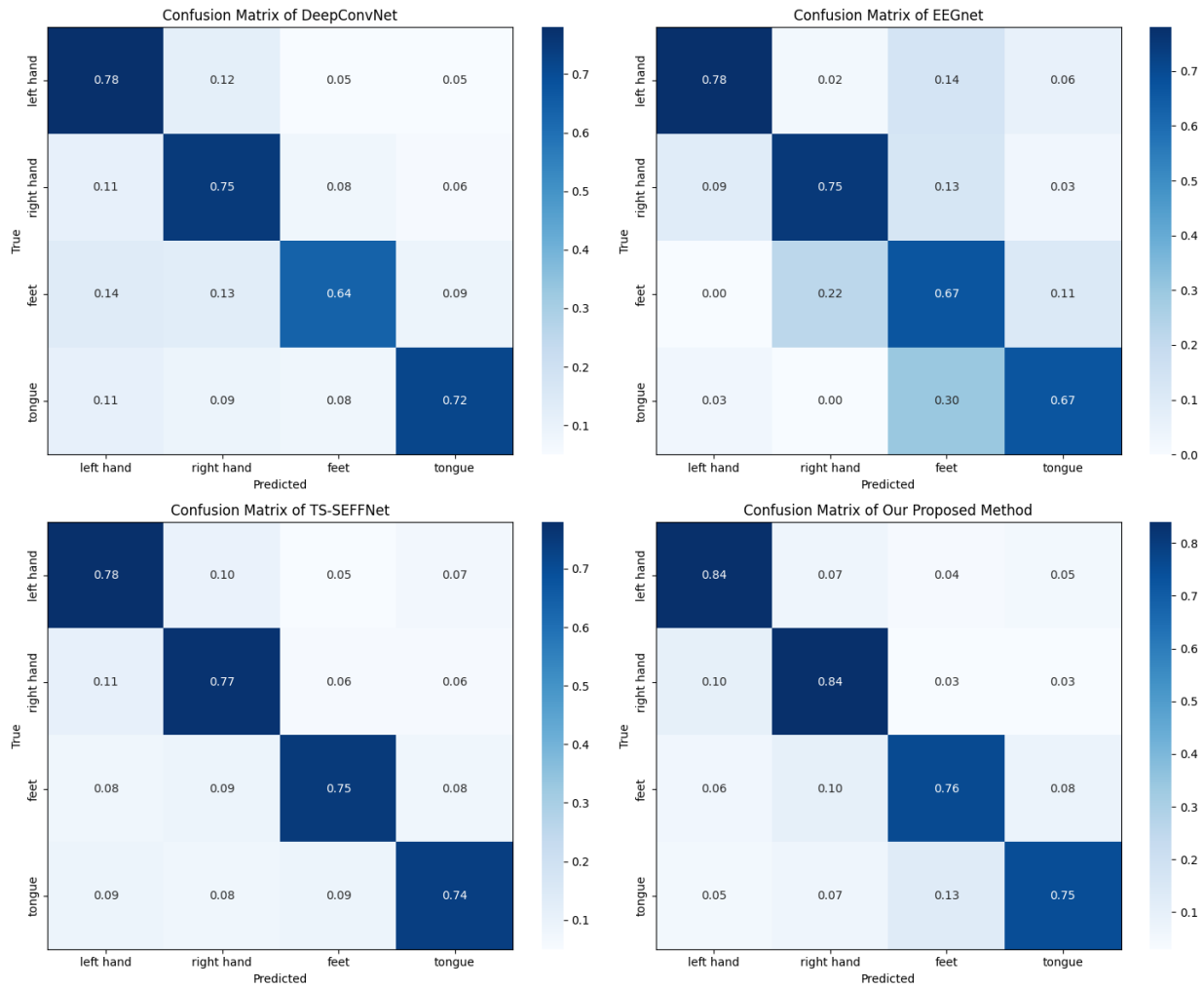


Fig. 6. Confusion matrix of the methods.

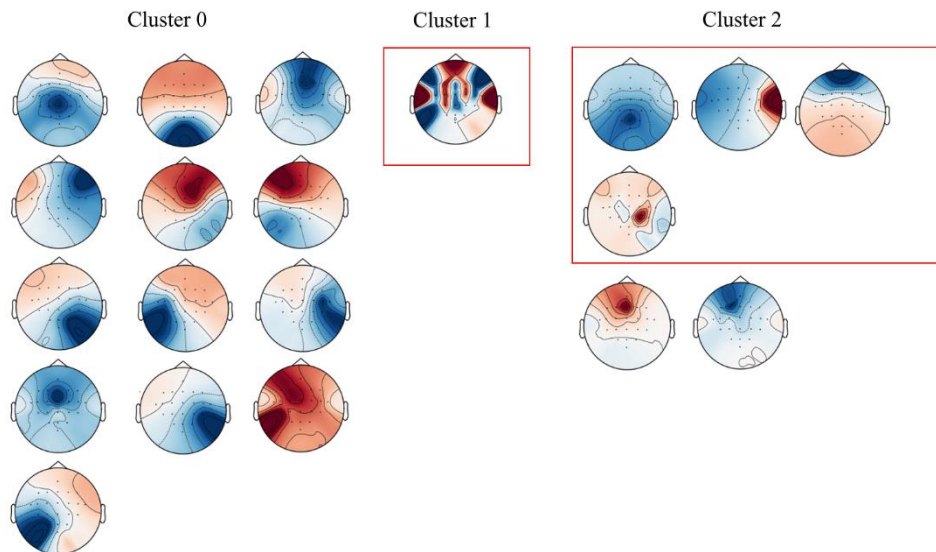


Fig. 7. Best clustering result for subject 1.



#### IV. DISCUSSION

From the clustering results in Fig. 7, it can be seen that the combination of features (kurtosis, skewness, and covariance) can effectively cluster brain state components and non-brain state components into different clusters. The brain states within the red box are labeled as artifacts. However, in Cluster 2, two brain state components are still clustered together with artifacts. From the characteristics of these brain states, except for the central region of activation, the other regions of these two components are relatively inactive, that is similar to other artifact components in this category. This suggests that these two components may have some similarity in their features, resulting in clustering with artifacts. In the ablation experiment in Table I, the method of clustering and artifact removal using kurtosis, skewness and covariance features still achieves good average accuracy compared to not using this method. This proves that the selected features are still applicable to the remaining subjects. Compared to manually removing artifacts from nine subjects' ICs, the ICA+K-means method eliminates the time-consuming process of manually screening and removing artifact components for all subjects while achieving a performance loss of less than 1% in average accuracy. This is crucial for large datasets with many subjects and sessions. Manual identification and removal of ICs for each subject would be time-consuming and inefficient in large datasets. On the other hand, the automated process of the ICA+K-means method can quickly and accurately remove artifacts, saving significant labor and time costs.

As shown in Fig. 8, this work also performed a feature visualization of the comparison experiment. This was done by performing parameter extraction prior to the final classification of each network and then visualizing it via the t-SNE method. Different colors represent different parts of the motor imagery being performed: green for the left hand, purple for the right hand, blue for the tongue, and red for the feet. Different colors are used to distinguish the visualization results in order to represent them more intuitively. The visualization results of the proposed method have a smaller intra-class spacing than the other methods. This is consistent with the previous experimental results in Table III.

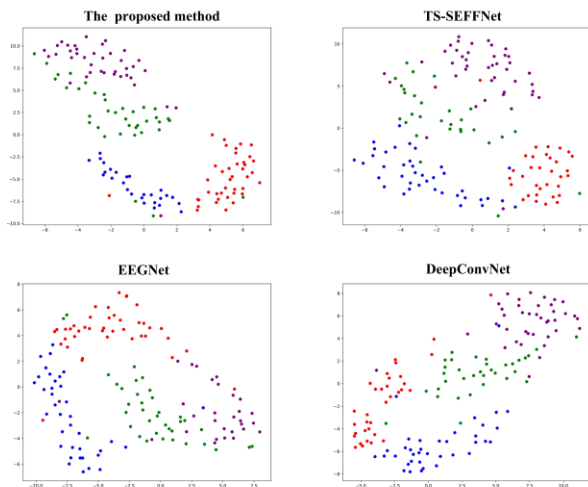


Fig. 8. Visualization results of comparison experiments.

EEG signals primarily require the extraction of temporal and spatial information. Previous studies have demonstrated that different people show specific electrical signal patterns in a certain range of frequency bands when imagining the same motor imagery task. However, the specific frequency bands may vary across individuals. Therefore, similar to the FBCSP approach, this study employs filters at multiple scales to capture temporal information, aiming to mitigate the impact of individual differences on temporal information extraction. For spatial information extraction, instead of electrode selection employed in some studies [30-34], the proposed method utilizes a spatial attention strategy after performing spatial convolution to automatically allocate weights to feature maps. The reason for this choice is that the dataset itself only has 22 EEG channels and performing electrode selection would result in the removal of some channels, which could have a negative impact on the result, especially in datasets with limited channels. To make full use of the feature maps, the proposed method utilizes spatial attention to automatically allocate weights to them. The results of the ablation experiment in Table III confirm the effectiveness of the multi-scale block and spatial block.

In addition to the above high-performance models, this work has also investigated some methods from recent MI-BCI studies that use the same dataset. Wang et al. [35] proposed an unsupervised domain adaptation framework called Iterative Self-training Multisubject Domain Adaptation (ISMDA) for the offline MI task, achieving an average classification accuracy of 69.51%. Liu et al. [36] proposed a SincNet-based hybrid neural network (SHNN) for MI-based BCIs to improve information utilization, achieving an average classification accuracy of 74.26%. She et al. [37] proposed an improved domain adaptation network based on Wasserstein distance, which utilizes existing labeled data from multiple subjects (source domain) to improve the performance of MI classification on a single subject (target domain), achieving an average classification accuracy of 77.6%. Fang et al. [38] proposed a fusion method combining Filter Banks and Riemannian Tangent Space (FBRTS) in multiple time windows to obtain more robust features, achieving an average classification accuracy of 77.7%. The comparative results are shown in Table IV.

TABLE IV. COMPARISON OF RECENT STUDIES

Method	Acc
ISMDA <sup>[35]</sup>	69.51%
SHNN <sup>[36]</sup>	74.26%
domain adaption network based on Wasserstein distance <sup>[37]</sup>	77.6%
FBRTS <sup>[38]</sup>	77.7%
Ours	79.83%

The optimal best feature combination used in this study achieved high performance on the BCI-IV-2a dataset. However, different datasets may have different parameters such as the number of electrodes, the number of subjects and the task types [39-41]. Therefore, when applying the proposed method to other publicly available datasets, parameters used in the experiment, such as the optimal feature combination, the

number of training epochs for the network and the number of clusters for clustering should be adaptively adjusted further. Since deep neural networks require a large amount of data for training, some data augmentation or other methods may be required by the experimenter to avoid model overfitting if the method is to be reproduced on a smaller dataset. Moreover, the performance of the model can be further improved if more useful features are provided and optimal parameters are searched. However, as the number of features increases, combining and selecting them self-adaptively for a specific subject will be discussed in future work.

## V. CONCLUSION

This study proposes a multi-scale CNN with a novel artifact removal strategy and spatial attention module for motor imagery recognition. By appropriately combining the selected features, it automatically removes artifacts using clustering algorithms on the components extracted by ICA, while ensuring high classification accuracy. The multi-scale convolutional blocks in SA-MSCNN, composed of different kernel sizes, extract multi-scale semantic features from the raw EEG data for classification purposes. The feature maps are then refined using a spatial attention module. The dense layer obtains the final classification results. To validate the effectiveness of this framework, the model has been applied to the BCI Competition IV-2a dataset. Compared to other existing excellent algorithms, this algorithm shows a significant improvement in classification accuracy. Experimental results demonstrate that this algorithm achieves high classification accuracy with an average accuracy of 79.83%. The current framework exhibits good classification performance and generalization. Compared to widely used EEGNet and DeepConvNet, the average classification accuracy improves by 7.39% and 7.84%, respectively. Compared to the newer state-of-the-art TS-SEFFNet, it achieves average classification accuracy improvements of 5.12%. This work also compares it with other recently published methods and the result shows the competitiveness of the proposed method. The proposed model can extract more effective features from EEG signals. This work contributes a novel method for automatic EEG artifact removal and an effective deep-learning model. It can be used to design efficient and accurate MI-based brain-computer interface frameworks to assist individuals with disabilities.

## ACKNOWLEDGMENT

The work was supported by the Science and technology key project of Henan Province (232102210017).

## DATA AVAILABILITY STATEMENT

The data presented in this study are openly available at the following URL/DOI: <https://bbci.de/competition/iv/>.

## REFERENCES

- [1] S. Kotchetkov, B. Y. Hwang, G. Appelboom, C. P. Kellner and E. S. Connolly Jr. "Brain-computer interfaces: military, neurosurgical, and ethical perspective." *Neurosurgical Focus*, vol. 28 5, 2010, pp. E25.
- [2] Xiaoqian Mao, Mengfan Li, Wei Li, Linwei Niu, Bin Xian, Ming Zeng and Genshe Chen, "Progress in EEG-Based Brain Robot Interaction Systems." *Computational Intelligence and Neuroscience*, vol 2017, 2017.
- [3] D. T. Bundy, L. Souders, K. Baranyai, L. Leonard, G. Schalk, R. Coker, Daniel W Moran, T. Huskey and E. C. Leuthardt, "Contralesional brain-computer interface control of a powered exoskeleton for motor recovery in chronic stroke survivors," *Stroke*, vol. 48, no. 7, pp. 1908–1915, 2017.
- [4] A. D. Moldoveanu, O. Ferche, F. Moldoveanu, R. G. Lupu, D. Cinteza, D. C. Irimia and C. Toader, "The TRAVEE system for a multimodal neuromotor rehabilitation," *IEEE Access*, vol. 7, pp. 8151–8171, 2019.
- [5] M. Staffa, M. Giordano, and F. Ficuciello, "A wisard network approach for a bci-based robotic prosthetic control," *International Journal of Social Robotics*, vol. 12, pp. 749–764, 2020.
- [6] R. Mane, T. Chouhan, and C. Guan, "BCI for stroke rehabilitation: motor and beyond," *Journal of Neural Engineering*, vol. 17, no. 4, p. 041001, aug 2020.
- [7] M. Sebastián-Romagosa, W. Cho, R. Ortner, N. Murovec, T. V. Oertzen, K. Kamada, B. Z. Allison and C. Guger, "Brain computer interface treatment for motor rehabilitation of upper extremity of stroke patients—A feasibility study," *Frontiers in Neuroscience*, vol. 14, pp. 1–12, Oct. 2020.
- [8] A. Biasiucci, B. Franceschiello, M. M. Murray, "Electroencephalography." *Current Biology*, vol 29, 2019, pp. R80-R85.
- [9] A. Mognon, J. Jovicich, L. Bruzzone, and M. Buiatti, "Adjust: An automatic eeg artifact detector based on the joint use of spatial and temporal features," *Psychophysiology*, vol. 48, no. 2, pp. 229–240, 2011.
- [10] M. Chaumon, D. V. Bishop, and N. A. Busch, "A practical guide to the selection of independent components of the electroencephalogram for artifact correction," *Journal of Neuroscience Methods*, vol. 250, pp. 47–63, 2015.
- [11] A. Hyvärinen and E. Oja, "Independent component analysis: Algorithms and applications." *Neural Networks* vol.13, pp. 411-430, 2000.
- [12] V. D. Calhoun, Jingyu Liu, T. Adali. "A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data." *Neuroimage* vol. 45, 2009, pp. S163-S172.
- [13] I. Winkler, S. Haufe and M. Tangermann. "Automatic classification of artifactual ICA-components for artifact removal in EEG signals." *Behavioral and Brain Functions* vol. 7, 2011, pp. 1-15.
- [14] A. M. Judith, S. B. Priya and R. K. Mahendran. "Artifact Removal from EEG signals using Regenerative Multi-Dimensional Singular Value Decomposition and Independent Component Analysis." *Biomedical Signal Processing and Control* vol. 74, 2022, p. 103452.
- [15] H. Varsehi, S. M. P. Firoozabadi. "An EEG channel selection method for motor imagery based brain-computer interface and neurofeedback using Granger causality." *Neural Networks*, vol. 133, 2021, pp. 193-206.
- [16] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, B. J. Lance. "EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces." *Journal of Neural Engineering* vol. 15(5), 2018, p. 056013.
- [17] R. T. Schirrmester, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger and M. Tangermann, et al. "Deep learning with convolutional neural networks for EEG decoding and visualization." *Human Brain Mapping* vol. 38,11 (2017): 5391-5420.
- [18] Yang Li, Lianghui Guo, Yu Liu, Jingyu Liu and Fangang Meng. "A Temporal-Spectral-Based Squeeze-and-Excitation Feature Fusion Network for Motor Imagery EEG Decoding." *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, 2021, pp. 1534-1545.
- [19] A. Vaswani, N. M. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser and I. Polosukhin, "Attention is all you need." *Advances in Neural Information Processing Systems*. 2017.
- [20] S. Woo, J. Park, JY. Lee, IS. Kweon, "CBAM: Convolutional Block Attention Module." In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds) *Computer Vision – ECCV 2018*. ECCV 2018. Lecture Notes in Computer Science(), vol 11211. Springer, Cham.
- [21] A. Galassi, M. Lippi and P. Torroni, "Attention in Natural Language Processing," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 10, pp. 4291-4308, Oct. 2021.

- [22] G. A. Altuwajjri and G. Muhammad, "Electroencephalogram-Based Motor Imagery Signals Classification Using a Multi-Branch Convolutional Neural Network Model with Attention Blocks." *Bioengineering* (Basel, Switzerland), vol. 9(7), p. 323, 2022.
- [23] L. Liu, C. Shi and X. Wu, "Low Quality Samples Detection in Motor Imagery EEG Data by Combining Independent Component Analysis and Confident Learning," 2022 21st International Symposium on Communications and Information Technologies (ISCIT), Xi'an, China, 2022, pp. 269-274.
- [24] MH. Lee, OY. Kwon, YJ Kim, HK. Kim, YE. Lee and J. Williamson, et al. "EEG dataset and OpenBMI toolbox for three BCI paradigms: an investigation into BCI illiteracy." *GigaScience* vol. 8,5 (2019): giz002
- [25] Kai Keng Ang, Zhang Yang Chin, Haihong Zhang and Cuntai Guan. "Filter Bank Common Spatial Pattern (FBCSP) in Brain-Computer Interface," IEEE International Joint Conference on Neural Networks. IEEE, 2008.
- [26] C. Brunner, R. Leeb, G. Müller-Putz, A. Schlögl, and G. Pfurtscheller, "BCI competition 2008—Graz data set A," *Inst. Knowl. Discovery, Lab. Brain-Comput. Interfaces, Graz Univ. Technol., Graz, Austria, Tech. Rep.*, 2008, pp. 136–142.
- [27] G. Dornhege, J.D.R. Mill'án, T. Hinterberger, D. McFarland, K.R. Müller, *Toward brain-computer interfacing*, MIT press, Cambridge MA, 2007.
- [28] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier and C. Brodbeck, et al. "MEG and EEG data analysis with MNE-Python." *Frontiers in Neuroscience* vol. 7 267. 26 Dec. 2013.
- [29] A. Delorme and S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial eeg dynamics including independent component analysis," *Journal of Neuroscience Methods*, vol. 134, no. 1, pp. 9–21, 2004.
- [30] Jiazhen Hong, F. Shamsi and L. Najafizadeh. "A Deep Learning Framework Based on Dynamic Channel Selection for Early Classification of Left and Right Hand Motor Imagery Tasks." Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference vol. 2022 (2022): 3550-3553.
- [31] Z. A. A. Alyasseri, O. A. Alomari, S. N. Makhadmeh, S. Mirjalili, M. A. Al-Betar, S. Abdullah, et al., "EEG Channel Selection for Person Identification Using Binary Grey Wolf Optimizer," in *IEEE Access*, vol. 10, pp. 10500-10513, 2022.
- [32] Wei Mu, Tao Fang, Pengchao Wang, Junkongshuai Wang, Aiping Wang, Lan Niu, et al, "EEG Channel Selection Methods for Motor Imagery in Brain Computer Interface," 2022 10th International Winter Conference on Brain-Computer Interface (BCI), Gangwon-do, Korea, Republic of, 2022, pp. 1-6.
- [33] J. Wang, L. Shi, W. Wang and Z. -G. Hou, "Efficient Brain Decoding Based on Adaptive EEG Channel Selection and Transformation," in *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 6, no. 6, pp. 1314-1323, Dec. 2022.
- [34] Abdullah, I. Faye, and M. R. Islam, "EEG Channel Selection Techniques in Motor Imagery Applications: A Review and New Perspectives." *Bioengineering* (Basel, Switzerland), vol. 9(12), p. 726, 2022.
- [35] He Wang, Peiyin Chen, Meng Zhang, Jianbo Zhang, Xinlin Sun and Mengyu Li et al. "EEG-Based Motor Imagery Recognition Framework via Multisubject Dynamic Transfer and Iterative Self-Training." *IEEE Transactions on Neural Networks and Learning Systems*, vol. PP 10.1109/TNNLS.2023.3243339. 20 Feb. 2023.
- [36] Chang Liu, Jing Jin, Ian Daly, Shurui Li, Hao Sun and Yitao Huang et al. "SincNet-Based Hybrid Neural Network for Motor Imagery EEG Decoding." *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30 (2022): 540-549.
- [37] Qingshan She, Tie Chen, Feng Fang, Jianhai Zhang, Yunyuan Gao and Yingchun Zhang. "Improved Domain Adaptation Network Based on Wasserstein Distance for Motor Imagery EEG Classification." *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. PP 10.1109/TNSRE.2023.3241846. 1 Feb. 2023.
- [38] Hua Fang, Jing Jin, Ian Daly and Xingyu Wang. "Feature Extraction Method Based on Filter Banks and Riemannian Tangent Space in Motor-Imagery BCI." *IEEE Journal of Biomedical and Health Informatics*, vol. 26,6 (2022): 2504-2514.
- [39] Hohyun Cho, Minkyu Ahn, Sangtae Ahn, Moonyoung Kwon and Sung Chan Jun. (2017). "EEG datasets for motor imagery brain-computer interface." *GigaScience*, vol. 6(7), pp. 1–8.
- [40] Jun Ma, Banghua Yang, Wenzheng Qiu, Yunzhe Li, Shouwei Gao and Xinxing Xia. (2022). "A large EEG dataset for studying cross-session variability in motor imagery brain-computer interface." *Scientific Data*, vol. 9(1), p. 531.
- [41] J. Shin, A. Luhmann, B. Blankertz, DW. Kim, J. Jeong, HJ. Hwang, et al. "Open Access Dataset for EEG+NIRS Single-Trial Classification." *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. 2017; vol. 25(10), pp. 1735-1745.

# SFFT-CapsNet: Stacked Fast Fourier Transform for Retina Optical Coherence Tomography Image Classification using Capsule Network

Michael Opoku<sup>1</sup>, Benjamin Asubam Weyori<sup>2</sup>, Adebayo Felix Adekoya<sup>3</sup>, Kwabena Adu<sup>4</sup>

Department of Computer Science and Informatics, University of Energy and Natural Resources, Sunyani, Ghana<sup>1,4</sup>  
Department of Electrical and Computer Engineering, University of Energy and Natural Resources, Sunyani, Ghana<sup>2</sup>  
Faculty of Computing, Engineering, and Mathematical Sciences, Catholic University of Ghana, Sunyani<sup>3</sup>

**Abstract**—The work of the Ophthalmologist in manually detecting specific eye related disease is challenging especially screening through large volume of dataset. Deep learning models can leverage on medical imaging like the retina Optical Coherence Tomography (OCT) image dataset to help with the classification task. As a result, many solutions have been proposed based on deep learning-based convolutional neural networks (CNNs). However, the limitations such as inability to recognize pose, the pooling operations which affect resolution of the featured maps have affected its performance in achieving the best accuracies. The study proposes a Capsule network (CapsNet) with contrast limited adaptive histogram equalization (CLAHE) and Fast Fourier transform (FFT), a method we called Stacked Fast Fourier Transform-CapsNet (SFFT-CapsNet). The SFFT was used as an enhancement layer to reduce noise in the retina OCT image. A two-block framework of three-layer convolutional capsule network each was designed. The dataset used for this study was presented by University of California San Diego (UCSD). The dataset consists of 84,495 X-Ray images categorized into four classes (NORMAL, CNV, DME, and DRUSEN). Experiment was conducted on the SFFT-CapsNet model and results were compared with baseline models for performance evaluation using accuracy, sensitivity, precision, specificity, and AUC as evaluation metrics. The evaluation results indicate that the proposed model outperformed the baseline model and state-of-the-arts models by achieving the best accuracies of 99.0%, 100%, and 99.8% on overall accuracy (OA), overall sensitivity (OS), and overall precision (OP), respectively. The result shows that the proposed method can be adopted to aid Ophthalmologist in retina disease diagnosis.

**Keywords**—Capsule network; convolution neural network; medical imaging; optical coherence tomography

## I. INTRODUCTION

The concept of identifying specific medical conditions in the human body through the analysis of medical images to establish basis for the existence and growth rate of a particular disease can be very tedious and stressful. As a radiologist, one is confronted with the burden of finding and interpreting extracted features from medical images to diagnose and monitor different kinds of diseases associated with human body [1-3]. Human beings in our nature are not only slow in processing medical images but are prone to errors especially when stressed out. Considering a sensitive area like medical field, a wrong diagnose can be quite expensive as its

implications can lead to unexpected consequence [4]. As a result, the attention of many researchers has been diverted into finding a substantive solution to assist the radiologists to deal with the complications confronted on daily bases as part of their work [1]. Computer aided models can help but the major challenge has been selecting the right method and acquiring good performance such as high prediction accuracy, achieving low runtime and low computational cost [5-12].

The introduction of deep learning (DL) brought a promising breakthrough especially in the field of medical science. The Artificial Neural Networks (ANN) [13-14] and Convolutional Neural Networks (CNN) [15] which are DL techniques became the most predominantly employed methods in identifying and diagnosing anomalies using radiological imaging technologies. However, both the ANN and the CNN have their benefits and limitations which usually affect performance of the model. They both require huge dataset for efficient training of models and increasing the image resolution adds up to number of trainable parameters which affects runtime and computational cost of the model [16]. According to Noord, and Postma, [17] ANN is not flexible and does not allow for easy customization of the model. Moreover, ANN also has deficiency with diminishing and exploding of gradient [17]. The CNNs framework gained more attention as a result of the high performance it can offer and its flexibility to use. The enormous computer-aided algorithms providing state-of-the-art early disease detection results for the medical imaging diagnosing tasks have depended on the building blocks of CNNs for most of the models produced [18].

However, the million unanswered yet important question is whether CNN models truly generalize. According to Gu, [19], a well-trained CNN model is still prone to adversarial attack as the network can easily be fooled by images that are carefully designed with imperceptible perturbations. According to Xi et al., [20], the CNNs face two major challenges which are lack of rotational invariance and failure to consider the spatial orientation hierarchies between features of the image. CNNs therefore, require huge dataset with different poses for same images, if one wants to achieve high performance for such classification model.

In an attempt to address these challenges, Hinton et al., [21] introduced novel type of neural network known as the capsule network (CapsNet). Sabour et al., [22] enhanced the CapsNet

architecture by introducing the dynamic routing by agreement between capsules. The CapsNet used the routing by agreement to resolve the problem resulting from the pooling operations of the CNNs which affects the resolution of the feature maps. Again, the CapsNet employed what is known as the reconstruction regularization to recognize the spatial hierarchical relationships among the entity parts. Moreover, since the CapsNet is equivariant in nature, it does not require huge dataset or data augmentation like rotation and scaling during training and testing also, adversaries' attacks such as the imperceptible pixel perturbation of the CNNs is also addressed by the CapsNet.

The study therefore adopts the CapsNet to classify retina optical coherence tomography (OCT) images due to the many advantages it offers over the other DL models. The study reconstructs the CapsNet architecture to include other controlled parameters to enhance its performance significantly. To ensure effective distribution of coupling coefficient which can enhance performance accuracy and convergence, the study performs normalization using the sigmoid function [23] instead of the SoftMax [22]. The study makes the following contributions:

- Proposes new capsule networks architecture named Stacked Fast Fourier Transform capsule network (SFFT-CapsNet).
- Evaluate the proposed model on retina OCT dataset.
- Compares results of proposed model with original capsule networks and state-of-the-art deep learning models performance.
- Provides visualization of internal processing results to establish better explainability of the proposed architecture.

The rest of the study is organized in the following manner;

The Section II of the study provides insight on related works. The Section III provides the methods of the study. The Section IV presents the experimental setup and results discussion. Finally, the Section V presents conclusion and recommendation for future expansion.

## II. RELATED WORKS

Every computer vision has a simple task of performing classification on given images or objects. Deep learning models have been applied to different domains like medical field for classification task. Wang et al., [24] implemented a deep learning model for automatic detection of metastatic breast cancer. The method employed enhanced the localization task. Nithya et al., [25] implemented ANN to detect kidney disease like kidney stones. Arunkumar et al. [26] implemented another algorithm based on ANN to classify the abnormal magnetic resonance (MRI) image which was used to identify brain tumor.

According to Karri et al., [27], transfer learning was easily included in CNN architecture to train on small dataset after which the results were successfully implemented on OCT image for classification of DME and dry AMD. However, due

to the limitation of the CNNs mentioned earlier, the CapsNet [22] was introduced to provide better classification which also addressed most of the failures of the CNNs.

The performance of CapsNet has always been compared to CNNs [28] in many researches to help enhance the algorithm and architecture of the CapsNet. The results of the comparison also indicate that CapsNet also has its own challenges [29]. A major challenge identified by Sabour et al., [22] indicated that the CapsNet architecture attempted to extract features on every entity part found of the image which might not be necessary for classification task. Extracted information such as background information makes CapsNet implementation more vulnerable to misclassification [21]. The explanation given to this was that the shallow structure of the CapsNet implemented with single convolutional layer is not sufficient enough to extract only required features. Liu et al., [30] demonstrated in their article that the original CapsNet performed very poor on complex data due to insufficient convolutional layers required to extract better features to enhance performance. This conclusion was made when the original CapsNet was compared with their proposed model called the DDRM-CapsNet for performance efficiency. In their study, they modified the original CapsNet architecture to include more convolutional layers to ensure better feature extraction. The study also enhanced the dynamic routing mechanism to include two stages and changed the output vector to 24 dimensions. So, in all, the network had three standard convolutional layers, a primary capsule layer, two-digit capsule layers and three fully connected layers.

To improve the architecture of the CapsNet, many researchers attempted improving the algorithm to include new features. In the CapsNet architecture, information of each capsule is sent to the next available layers at the full magnitude of its activation value but still lacks an appropriate control mechanism for selecting discriminant features from the outputs of each layer [29]. This means in the routing mechanism, encoded information from one capsule to the next capsule layer is not filtered even if it contains background information or unwanted information that is not required for classification decision. Also, CapsNet introduced an initial logit  $b_{ij}$  in the routing Softmax function to represent the log prior probability of how tight the coupling of the initial capsule  $i$  is with capsule  $j$  which depended on location and type of two capsules. Nonetheless, the function instead transformed the logits of the coupling coefficients to what is known as concentrative values. This means background information can mistakenly be sent to the next capsule layer with very high coefficient that can impact large values for summation of the predicting vectors [23]. Hence, Zhao et al., [31] concluded in their article that the SoftMax function implemented for normalization process to ensure uniform assignment of probability values between capsules prevented fair distribution of coupling coefficients which affected the performance. In another research, Yang & Wang, [29] also proposed two different methods to address the issue of capsules obtaining high activation values. Their study also introduced Cubic-Increase Squash (CI – squash) and Powered Activation (PA) which was modification to the original version of the squash function. The study demonstrated information sensitiveness was a major reason why CapsNet was not achieving high performance with color

background images. Again, the study concluded there was the need to restrict the capsules from achieving unreasonably high distribution of activation values as it also affected the performance. As a result, Mensah et al, [32] implemented the power squash function as the original squash is susceptible to generating high activation values. The activation values of the original capsules experienced faster growth at the initial stage leading to very high generated activation values. It was therefore important to include a sparsity which constrains the capsules from achieving such high activation values so that the capsules would be able to identify distinguishing features that are of greater interest to the classification process.

Many researchers have also tried to find different means to improve the performance of the CapsNet models to resolve the issue of the information sensitiveness and the high activation values. Some focused on reengineering the CapsNet architecture while others tried to improve the algorithm [32-35]. Nguyen and Ribeiro [36] reconstructed the vector CapsNet architecture to include more convolution layers and also varied the fully connected layers to leverage better filter input images for best feature extraction and image restoration. In another study, Xiang et al. [37] developed a multi-scale capsule network for classifying images which sorted to address the shortfalls of the original CapsNet architecture. The proposed multi-stage CapsNet included structural and semantic information on the first layer. Phaye et al., [38] enhanced extraction of the discriminative feature maps by introducing dense and diverse CapsNet. Their study replaced the convolutional layer with densely connected convolutional layer to improve the performance. In another studies by Huang et al., [39], to address the issue of vanishing gradient problem, the study introduced a dense connection between every layer. The results also indicated a significant improvement of the model. Other studies also tried to introduce residual connections in their attempt to address the vanishing gradient problem [40-41]. Bhamid & El-Sharkawy, [42] also implemented a CapsNet model which allowed the primary capsule to carry information at three different image scales. Their study dealt with complex data such as CIFAR10 which the model showed a significant improvement in the classification accuracy.

The capsule architecture has been applied to many different areas where image classification was a problem. Li et al. [43] implemented a capsule network model to recognize and monitor the growth rate of the rice crops through the images captured with unmanned aerial device. According to another article by Paoletti et al. [44], a CNN-Based-CapsNet was implemented in their study to classify remotely taken hyperspectral images. A similar study was also conducted by on hyperspectral image classification [45-46]. The overall performance also indicated a promising result than using the convolutional network. The implementation of CapsNet in the area of medical imaging has also shown very promising results. Adu et al, [47], implemented Dilated CapsNet in their research on Brain Tumor Classification Also, Afshar et al, [9] implement another CapsNet model for brain tumor classification. Koresh and Chacko, [48] also implemented CapsNet noiseless image classification algorithm which was used to classify corneal optical coherence tomography (OCT) dataset.

In our studies, we look at introducing two different enhancement layers to increase the information sensitivity to enhance the performance of the model. We also replace SoftMax activation function (AF) with sigmoid AF and reconstruct the original architecture to include six convolutional layers which were as a result of best performance from different modifications.

#### A. Optical Coherence Tomography Images

The Optical Coherence Tomography (OCT) is an accepted standard clinical practice diagnostic imaging technique for diagnosing retinal diseases. The results of this method provide an OCT image with possible features enough to provide sufficient visualization for detecting if there is a change in the retinal vessels from the imprint of the OCT film. The evaluation parameter is usually to identify if there is an increase or decrease in the retina layer [26]. The OCT is usually employed to acquire very high-resolution cross-sectional images from the retina. It uses low-coherence light to capture two and three-dimensional images from optical scattering media like biological tissue. The OCT can be used to capture large number of images through which the level of deterioration of the optic nerves can be determined with time.

The method is very easy to implement and does not involve any ionizing radiation [49]. This makes it possible to differentiate between the various layers of the retina so that specific diagnose can be established based on the measurement of the retina thickness to detect a retinal disease. The OCT now serve as a baseline retinal assessment and popular choice capturing retinal image for clinical practice before therapy session is initiated [50-51]. Ophthalmologist might depend on the results of the OCT image to select a choice of treatment for their patient. Therefore, emphasis on the increasing essential role of the OCT imaging cannot be overlooked. There are so many diseases that can be diagnosed using OCT imaging. Disease like diabetic retinopathy which is as a result of damage to blood vessels of retina, Macular pucker, Glaucoma, Macular hole, Age-related macular degeneration, Drusen, Central serous retinopathy, Macular edema, Vitreous traction, and Optic nerve abnormalities other macular and other related diseases are visible to OCT images. There are many eye diseases resulting from deteriorating of these retina cells. The focus of this study is centered on classification of three major Macular diseases. According to research age is a major risk factor associated with macular disease. The common diseases that may affect the healthiness of the macula are age-related macular degeneration [AMD], choroidal neovascularization (CNV) and diabetic macular edema (DME) [50] [52-53].

#### B. The CapsNet Architecture

The concept of CapsNet was first introduced in the Transforming Auto-Encoders in 2011 by Hinton et al., [21]. In their article, it was concluded that the CNN loose valuable information such as pose and spatial relationships among features maps due to the Max pool operation. As a result, the concept of the Transforming Auto-Encoders which used capsules to encode entities instead of neuron was introduced to keep the meaningful information that could influence the output of the classification task. Nonetheless, the concept did not receive any attention until the dynamic routing by

agreement using CapsNet [22] was introduced. The CapsNet uses capsules to encode entities such that each capsule is represented as activity vector to indicate an instantiation parameter (e.g. Texture, color, angle etc.) of a specific entity. The activity vector elements encode the properties that represent the entity and their activation vector indicates a probability of pose that an instantiation feature of an entity exist within its limited domain. This means the direction of the capsules indicates detailed characteristics of the features and the length of the capsule indicates its probability of existence of different features. The CapsNet architecture can be represented as an inverse computer graphic [22]. The implementation of the CapsNet in many research have always resulted in producing high accuracy [54-55].

In its operation, the CapsNet takes input vector from the lower capsule layer and multiply them by the weight matrices and a coupling coefficient. The CapsNet is seen as an equivariant in nature and therefore is able to recognize pose such as size, position and direction as well as varieties of features that have been randomly placed. In an article presented by Lenssen et al., [56], it was indicated that the CapsNet does not only represent equivariance of the characteristics of the entity type but can also signify invariance of the existence probability. The design of the CapsNet is basically meant for classification of images through feature extraction like that of CNN. This brands the CapsNet as an efficient platform when it comes to dealing with establishing spatial relationship and dealing with hierarchical data. The meaningful information is stored at different levels of capsules and the higher the levels the greater the information they can accumulate. The various levels can be seen as lower level which represent the primary capsule and the higher level which represent the digitCapsule. The capsules become activated when certain conditions are satisfied.

Through the dynamic routing process which implements routing by agreement mechanism, each active capsule must select another capsule from the next layer that the features captured can be passed to. The process ensures that the lower-level capsules agree on a feature before it is sent to the higher-level (digit Capsule). Through the training process, each capsule is able to capture certain features or characteristics of the image and the assumption is that the capsule is activated if there are properties of the image required by the higher layer for which the capsule must respond. The routing process is a major feature in the CapsNet. Fig. 1 shows the Dynamic routing process. The dynamic routing process is implemented to update weights between capsules from one layer to next which allows characteristics or properties captured by lower node capsules to be propagated to the next suitable capsule at the upper layer. If the prediction matches the higher-level capsule's output, then the coupling coefficient for these two capsules is increased. Let  $u_i$  be the output of capsule  $i$  and its prediction from parent capsule  $j$  is expressed as:

$$S_j = \sum_i c_{ij} \hat{u}_{j|i} \quad (1)$$

A nonlinear function is used to shrink long and short vectors to 1 and 0 respectively. Eq. (2) shows the non-linear squash function.

$$v_j = \frac{\|s_j\|^2 s_j}{1 + \|s_j\|^2 \|s_j\|} \quad (2)$$

where  $s_j$  in Eq. (6) is the input vector to the  $j$ th capsule and  $v_j$  is the output vector. CapsNet adopts non-linearity squashing function on output vectors ( $v_j$ ) in each iteration [11]. This shows the likelihood of the vector between 0 and 1, which means that it squashes small vectors and maintains long vectors in the unit length.

$$\{v_j \approx \|s_j\| s_j = 0 \quad v_j \approx \frac{\|s_j\|}{\|s_j\|} \quad (3)$$

The log probabilities are updated in the routing process based on the agreement between  $v_j$  for the fact that the agreement between two vectors will be increased and have a large inner product. Therefore, agreement  $a_{ij}$  for updating the log probability and coupling coefficient is defined as:

$$a_{ij} = v_j \hat{u}_{j|i} \quad (4)$$

Capsule  $k$  in the last layer is connected with a loss  $l_k$ . This puts a big loss value on capsules with long output instantiation parameters when the entity does not exist. The loss function  $l_k$  is expressed as follows:

$$l_k = T_k \max(0, m^+ - \|v_k\|)^2 + \lambda(1 - T_k) \max(0, \|v_k\| - m^-)^2 \quad (5)$$

where  $T_k$  is 1 when class  $k$  is present, and is 0 otherwise. The  $m^+$ ,  $m^-$ , and  $\lambda$  are hyperparameters that are set before the learning process.

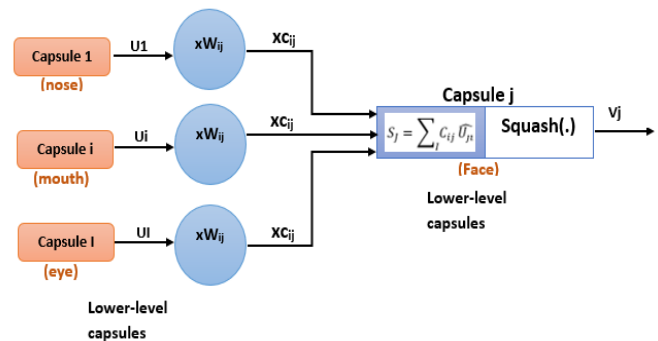


Fig. 1. Dynamic routing process.

### III. PROPOSED METHOD

The main objective of the study is to design a capsule network model that include image processing tools to ensure textural enhancement for better feature extraction to improve performance. Increasing the depth of the model too much can lead to overfitting, so the study carefully introduced two blocks of three convolutional layers through several model modifications to address the insufficiency nature of the convolutional layer in the original architecture. This was implemented while ensuring the model does not become too complex which can lead to overfitting. The study employed two different enhancement techniques. The network after reconstruction makes the model more sensitive to input image and still suppresses overfitting as we introduce a dropout technique by setting the early-stopping hyperparameter (patience) to 10 epochs if validation loss does not improve

during training to save only best models [59]. After exploring several model modifications, we arrived at the following conclusions:

- Implementation of contrast limited adaptive histogram equalization (CLAHE) technique and Fast Fourier Transform (FFT) enhancement to reduce noise in the various input images for better textural feature extraction while reducing number of trainable parameters.
- Power Squash: The study adopted the power version of the original squash function  $\|V_j\|^n \frac{V_j}{\|v_j\|}$  based on [30] [32]. The power squash is able to suppress smaller activation values (see Fig. 2).
- Sigmoid Activation: The sigmoid [23] activation function improved the distribution of the coupling coefficients leading to the overall improvement of the model performance. Even though the original CapsNet used the SoftMax function which was believed to constrain the  $C_{ij}$  within a smaller interval [31]. The goal is to acquire a coefficient that is sufficient enough to produce large values for better distribution. This way, it will be able to establish relevant features that are required by prediction vectors. Based on experimental results obtained from the different model modifications, there is a clear indicated that sigmoid activation function improved the convergence and overall accuracy of the model.
- Loss Function: The study introduced a loss function called E-swish to enhance the performance of the model instead of using existing activation function. This was an enhanced version of the original swish function. We compare the performance of our activation function with RELU and from the experimental results our function outperformed the existing baseline activation functions.

Power Squash: The study adopted the power version of the original squash function based on [29] [32]. The power squash has capability to compress short vectors to almost zero length while extending the long vectors to a value slightly below one. However, the original squash function generated high activation values for smaller  $\|s_j\|$  which intend generated very high activation values that are not sufficient to maintain high information sensitivity for the CapsNet model. Therefore, sparsity is required to constrain the capsules from achieving high activation values. Sparsity as explained by [32] is employed to differentiate and consider highly discriminant capsules that can extract required information from complex images especially ones with varied backgrounds. Fig. 2 shows the original squash function versus the suppressed small values of the power squash function. The power squash however, introduced sparsity to the model by controlling how the initial activation values are computed by the primary capsule. Therefore, for high values of  $n$ , the function experience very slow values at the initial stage and appreciated as the  $\|s_j\|$  increased. This was not the case of the original squash function which experienced a sharp increase at the initial stage as indicated in the Fig. 2.

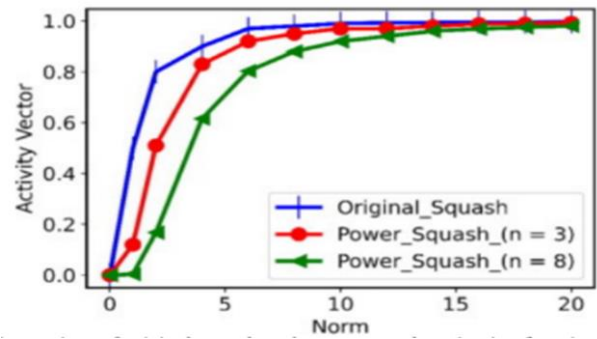


Fig. 2. The power squash versus the original squash.

#### A. Contrast limited Histogram Equalization and Fourier Transform

The study employed Contrast Limited Adaptive Histogram Equalization and Fourier Transform techniques to enhance the contrast of textural features and spatial patterns of the various input images. The conclusion for employing the two enhancements strategy was due to the results from the best performance of many model modifications. The method ensures visibility of the features for better feature extraction. The two enhancement methods have been deployed separately in many different researches due to their flexibility in implementation with low computation work load. The contrast limited adaptive histogram equalization (CLAHE) has mostly been deployed in research to reduce noise and improve the color of the x-ray images. It works best for all biomedical images by removing noise that can lead to miss classification. Histogram equalization produces over brightness of the input image which makes it difficult for the model to identify most required patterns and hidden objects.

Therefore, the CLAHE is employed to amplify the contrast of the image and limit neighboring pixel's procedure to reduce noise on the image. On the other hand, the Fourier Transform allow for images to be represented as a sum of complex exponentials of broad range frequencies. It has been implemented successfully in image processing applications such as enhancement, restoration or compression. It's usually employed as a processing tool to decompose images into its sine and cosine components. The output can then be represented in a Fourier or frequency domain when their input images are represented in a spatial domain equivalent. The Fourier domain allows each point to represent specific frequency in the spatial domain. It is best known for proving geometric characteristics of the spatial domain image as the images are decomposed into a sinusoidal component which makes it more flexible to analyze or process.

#### B. The Dataflow Analysis of the Model

Data augmentation was done by resizing the dataset images due to varied sizes of 1024x1050, 784x950, and 800x1020. Though the amount of retina OCT dataset used in this study was small however, CapsNet does not require huge dataset for training a model compared to CNNs. Fig. 3 shows summary diagram depicting the dataflow of the model.



### C. Model Architecture

The paper proposes a capsule network named Stacked Fast Fourier Transform CapsNet (SFFT-CapsNet). Fig. 4 illustrates the proposed Stacked Fast Fourier Transform CapsNet (SFFT-CapsNet) architecture. The SFFT-CapsNet consists of two blocks of convolutional layers. The convolutional layer block 1 consists of CLAHE layers, three convolutional layers, and batch normalization whereas the convolutional block 2 consists of Fourier transform layers, three convolutional layers, and batch normalization. The model consists of a Primary Capsule layer and a classification layer (retinaCaps). The process begins with passing all the input images with the dimension of  $48 \times 48 \times 3$  through CLAHE layer and the Fourier transform layer in the two blocks. The Fourier transform and CLAHE layer are image enhancement layers and therefore does not contribute additional parameters to the model. The output feature maps are forwarded to convolutional layers followed by batch normalization layers. After the input image goes through the enhancement layers, the output feature maps from the enhancement layer will still have the same dimension of  $48 \times 48 \times 3$  as the input image.

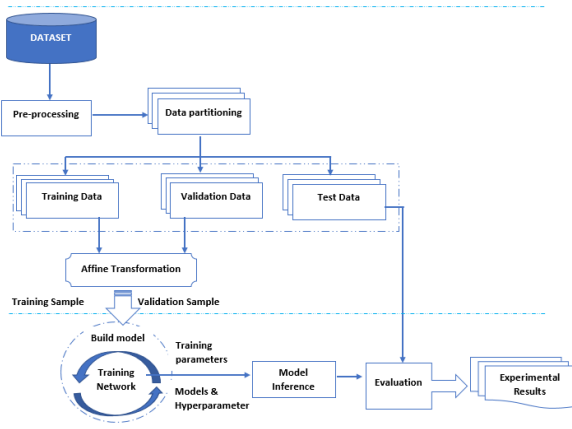


Fig. 3. Data flow diagram of the proposed model.

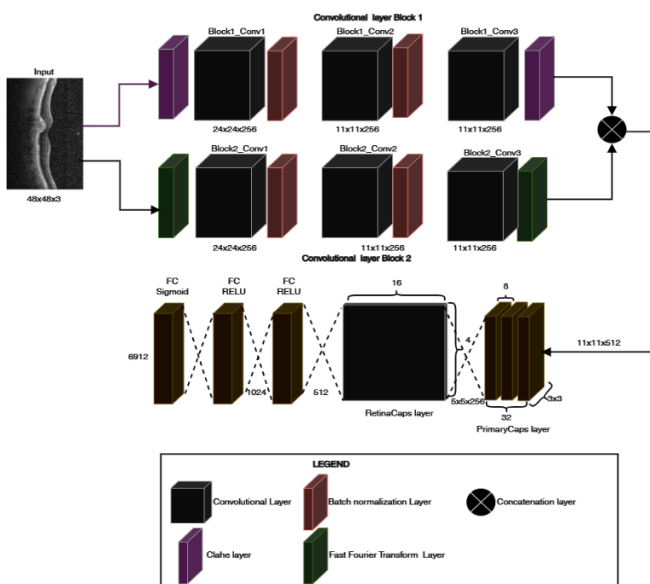


Fig. 4. Proposed Stacked Fast Fourier Transform CapsNet architecture.

The output  $48 \times 48 \times 3$  feature maps are sent to Block1\_Conv1 and Block2\_Conv1 with filter of 256, kernel size of  $3 \times 3$ , and stride of 2 which gives output feature map of  $24 \times 24 \times 256$ . These output feature maps are forward to the next convolutional layers which are Block1\_Conv2 and Block2\_Conv2 with the filters of 256, kernel sizes of  $3 \times 3$ , and strides of 2. This will convolve to feature maps of  $11 \times 11 \times 256$ . Again, these feature maps of  $11 \times 11 \times 256$  are sent to Block1\_Conv3 and Block2\_Conv3 with filters of 256, kernel sizes of  $1 \times 1$ , and strides of 1 to produce the feature maps of  $11 \times 11 \times 256$ . These feature maps from Block1\_conv3 and Block2\_Conv3 are concatenate to produce the feature map of  $11 \times 11 \times 512$  for the next layer. The feature maps from the feature extractions layers are forward to a PrimaryCaps with filter 256, kernel size of  $3 \times 3$ , and stride of 2 which will obtain output feature map of  $5 \times 5 \times 256$ . At the PC layer, a tensor product between  $u$  and the weights ( $W$ ) produces  $u \cdot w_j$  made up of 576 (i.e.,  $4 \times 4 \times 16$ ), 8-dimensional vectors. At the Digit Caps layer, the Recognition Caps will form  $k$ , 16D vectors, where  $k$  = number of classes. There are three fully connected (FC) layers in the decoder network consisting of 512, 1024, and 6912 neurons in the first, second, and third layers respectively.

### D. Experimental Settings

This practical aspect of the study was deployed using a Windows system with NVIDIA 394 GeForce GTX 1650 6GB GPU. The codes which used TensorFlow as the backend was implemented via Keras Libraries and python (Anaconda). Both the proposed CapsNet model (CLAHE-FT) and the original CapsNet were trained for 100 epochs respectively in order to compare their performance. The batch size of the input images was set to 32 while the learning rate was maintained at 0.0001. Through the deployment process, the study made use of the Adams algorithm with momentum as the gradient optimizer. The momentum was adjusted to 0.9 while the descent rate was set 10<sup>-6</sup>. To avoid overfitting as part of the reconstruction process, the early stopping hyperparameter thus patience was set to 10 during the training so the algorithm can only save the best model. To complement the reconstruction layer (FC) in avoiding overfitting, we set the early stopping hyperparameter; patience, to 10 during training and saved only the best model. The extracted code implemented is a modified code which is available at <https://github.com/XifengGuo/CapsNet-Keras>.

### E. Dataset

The dataset was made up of 84,495 x-ray images (jpeg) when it was downloaded from Kaggle.com. The folder contained subfolders which each contained specific category of images. In all, there were four categories of images which have been sampled into directories as Normal, CVN, DME and DRUSEN. However, the dataset had imbalance classes and since CapsNet could work with small size dataset, we decided to make it balanced for fair distribution by reducing the size of the various classes with large dataset. The images have been labeled as follows; (disease type)-(patient ID)-(image number of the patient). Again, these were OCT images that have been selected from retrospective cohorts of adult patients from institutions such as Shiley Eye Institute at University of California San Diego, California Retinal Research Foundation, Medical Center Ophthalmology Associates, Shanghai First 377

People’s Hospital, and Beijing Tongren Eye Center. It took barely four years to make such selections which were between the year 2013 and 2017. The inclusion criteria for eligible images were performed by different levels of trained and expert graders with enough experience to verify and establish correct data labels of the images into their respective classes. The first groups of graders were made up of undergraduate and medical students who had successfully passed an OCT interpretation course review.

These first graders were able to initiated quality control and excluded OCT images containing critical artifacts or significant image resolution reductions. Four ophthalmologists with a lot of experience were the second graders who independently graded the image that had passed the first grading. These second graders had a primary task of recording the present or absent of specific disease on each OCT scan. CNV, macular edema, drusen, and other pathologies which are present or absent on the OCT scan were recorded. The third group of graders consisted of two senior independent retinal specialists. Each specialist has over 20 years of clinical retinal experience, who varied the true label of the images. The Images that were imported into the database started with a label matching the recent diagnosis of the patient. The sample dataset selection is illustrated in a CONSORT-style as indicated in Fig. 5



Fig. 5. Sample retina OCT image representing different classes of images.

#### IV. EXPERIMENTAL RESULTS

This section presents the results of the proposed model used on the retina OCT image dataset. The SFFT-CapsNet model was established based on different modifications. We then compare the results to the original CapsNet by Sabour et al., [22] which has been serving as the baseline for most models in CapsNet. We evaluated the model by conducting comparative analysis on performance-accuracy with the current state-of-the-art models which have compelling efficient results on the same retina OCT dataset. This comparison is conducted to establish the best model for classification of the retina OCT images.

The study also establishes the efficiency of proposed model and its ability to generalize by comparing and visualizing the clusters formed through the routing process. An evaluation matrix such as accuracy (ACC), sensitivity (SE), precision (PR), specificity (SP), confusion matrix, receiver operating characteristic-area under ROC curve (ROC-AUC) are used to examine the performance of the models. The Precision and Recall were used to evaluate the model in order to achieve the Receiver Operating Characteristics (ROC) as well as the Precision Recall curves. The overall accuracy (OA), overall sensitivity (OS) and overall precision (OP) are also calculated. The computation of the OA for instance is by finding the average of the total accuracy scores for the four classes (CNV, DME, DRUSEN, NORMAL) from Table I as indicated in Eq.(6).

Thus,

$$OA = \frac{\text{correct classified classes}}{\text{total number of classes}} \tag{6}$$

The computation of the OS for instance is by finding the average of the total sensitivity scores for the four classes (CNV, DME, DRUSEN, NORMAL) from Table I as indicated in Eq. (7).

$$OS = \frac{\text{total sensitivity for the four classes}}{\text{total number of classes}} \tag{7}$$

The computation of the OP for instance is by finding the average of the total Precision-recall scores for the four classes (CNV, DME, DRUSEN, NORMAL) from Table I as indicated in Eq. (8).

$$OP = \frac{\text{total precision-recall for the four classes}}{\text{total number of classes}} \tag{8}$$

The results of the comparison between the original CapsNet and the SFFT-CapsNet is presented in Table I. The results indicated that the SFFT-CapsNet obtained overall accuracy of 99.0% which establishes a very high performance of the model over the original CapsNet which had an overall accuracy of 94.2% when applied to the retina OCT images. The performance of the model also indicated an overall sensitivity (OS) of 100% and overall precision of 99.8% for the SFFT as against the original CapsNet which had 94.5% and 97.0% respectively. Fig. 6 shows a histogram representing the comparison of accuracies based on the overall accuracies of OA, OS, and OP. Fig. 7 shows the validation and loss accuracy. Also, Fig 8 and 9 shows the Confusion Matrix and the ROC-AUC, respectively. Fig. 10 shows the Precision and Recall curves as applied on the dataset.

TABLE I. COMPARISON OF RESULTS OF THE SFFT-CAPSNET MODEL AND ORIGINAL CAPSNET

Method	Classes	ACC (%)	SE (%)	PR (%)	SP (%)	AUC (%)	OA (%)	OS (%)	OP (%)
Original CapsNet [22]	CNV	95.5	97.5	99.0	100	100	94.2	94.5	97.0
	DME	95.0	94.3	97.0	95.8	99			
	DRUSEN	98.8	88.2	96.0	97.6	99			
	NORMAL	87.6	98.2	96.0	98.2	97			
SFFT CapsNet [Ours]	CNV	100	100	100	100	100	99.0	100	99.8
	DME	98.8	100	100	100	100			
	DRUSEN	100	97.2	100	100	100			
	NORMAL	97.1	98.7	99.0	100	100			

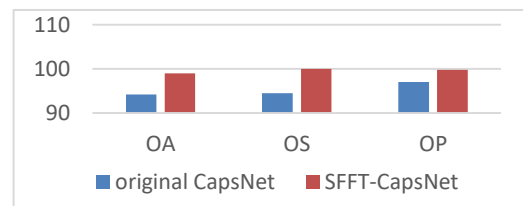


Fig. 6. Histogram comparing accuracies based on the overall accuracies of OA, OS, and OP.

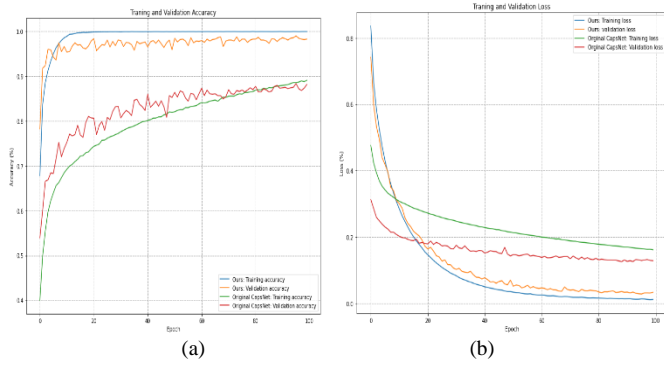


Fig. 7. Training, validation accuracy and loss curve on retina OCT images. (a) Training and validation accuracy curves of SFFT-CapsNet and original CapsNet., and (b) Training and validation loss curves of SFFT-CapsNet and original CapsNet.

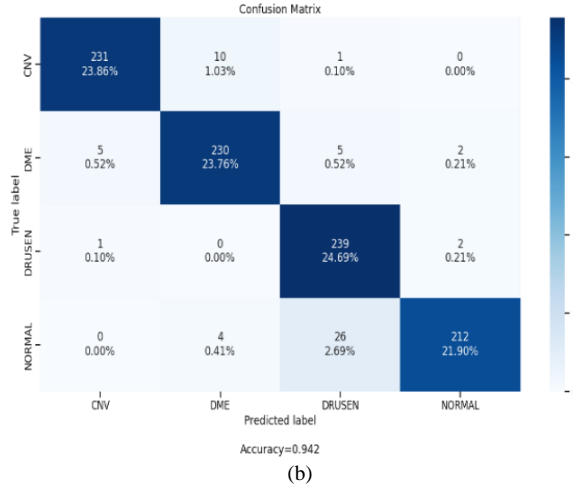
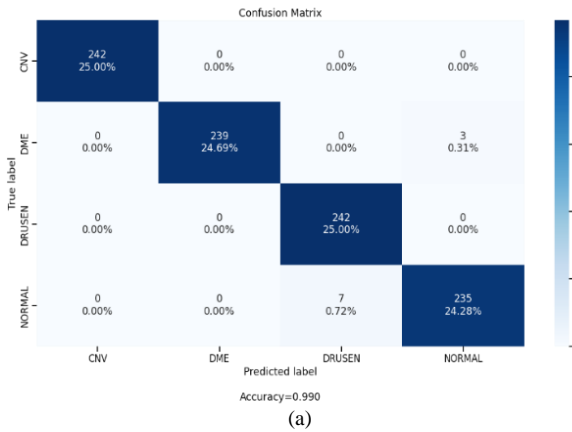


Fig. 8. Comparison of the proposed SFFT-CapsNet model and original CapsNet based on confusion matrix. (a) SFFT-CapsNet and (b) Original CapsNet.

A. Ablation Study

For every model, it is imperative to identify the various components which impacted significantly in the performance of the model. This can help to establish the robustness of our method and the various layers that contributed to improve the performance. To determine these components which impacted on the validation accuracy, an adjustment is made to the proposed architecture and its hyperparameters through several

experimental modifications to find the level of impact each component makes to the model. The results are then recorded and presented for further analysis. Table II shows the results of the ablation on retina OCT dataset. From the Table II, combination of the block-1 and the block-2 layers gave the best accuracy of 99.0% on the retina OCT dataset.

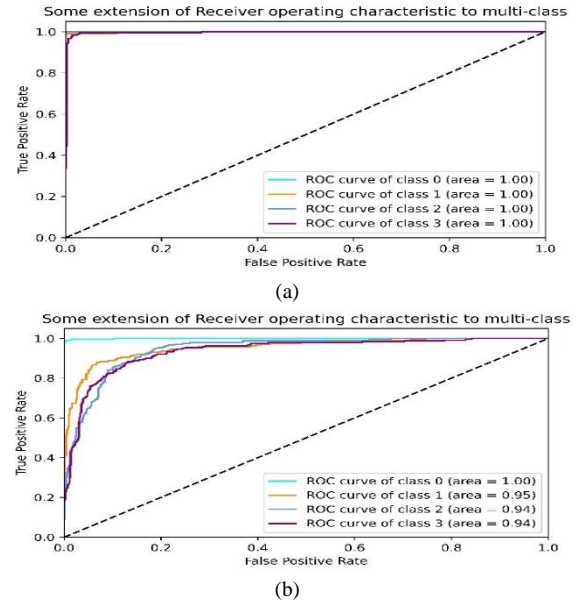


Fig. 9. Comparison of ROC-AUC on the proposed model and original CapsNet. (a) SFFT-CapsNet ROC (b) Original CapsNet ROC curve.

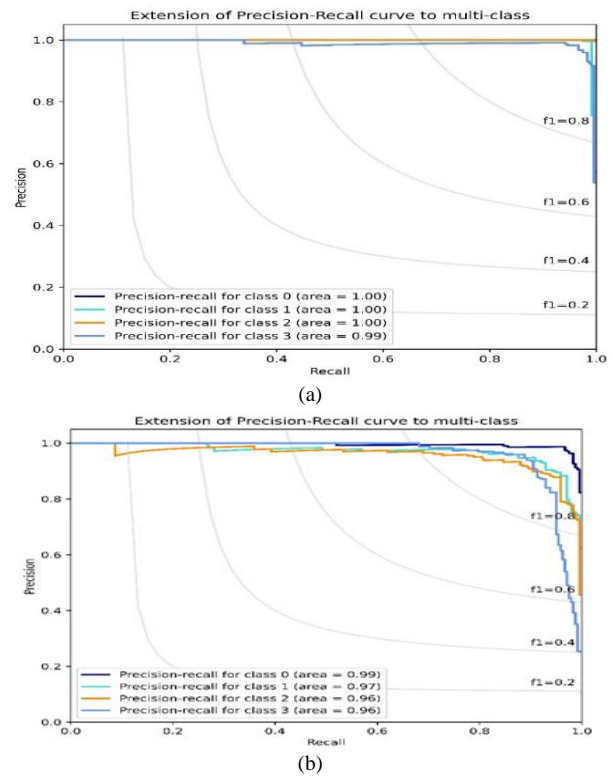


Fig. 10. Precision-recall curves comparison on the SFFT-CapsNet model and original CapsNet. (a) represents the SFFT-CapsNet Precision-Recall curve, and (b) Original CapsNet Precision-Recall curve.

TABLE II. RESULTS OF ABLATION STUDY ON SFFT-CAPSNET

No.	Layers	Number of Layers	Normalizer	Validation accuracy
1	Block-1(Conv layer)	1	SoftMax	94.22%
2	Block-1(Conv layer)	2	SoftMax	94.83%
3	Block-1(Conv layer)	3	SoftMax	95.02%
4	Block-1(Conv layer)	3	Sigmoid	95.91%
5	Block-1 (CLAHE layer, Conv layer)	1, 3	SoftMax	96.03%
6	Block-1(CLAHE layer, Conv layer)	2, 3	SoftMax	96.85%
7	Block-1(CLAHE layer, Conv layer)	2, 3	Sigmoid	97.70%
8	Block-2(FT,Conv layer)	1, 3	SoftMax	96.35%
9	Block-2(FT,Conv layer)	2, 3	SoftMax	97.52%
10	Block-2(FT, Conv layer)	2, 3	Sigmoid	98.01%
11	Block-1(CLAHE layer, Conv layer), Block-2(FT, Conv layer)	2,3, 2,3	SoftMax	98.71%
12	Block-1(CLAHE layer, Conv layer), Block-2(FT,Conv layer)	2,3, 2,3	Sigmoid	99.0%

A further analysis and evaluation were conducted by comparing our proposed SFFT-CapsNet with state-of-the-art results from other models that made use of the same retina OCT dataset. The evaluation matrix was based on accuracy, sensitivity, precision, specificity, overall accuracy, overall sensitivity, and overall precision. Table III presents the results of comparison of SFFT-CapsNet and previous works. The comparison was done considering the performance of the models on the individual classes of the retina OCT dataset. The letter “x” used in the table shows areas where the research paper failed to report the expected results for a particular evaluation metrics.

From the Table III, the best results have been highlighted in bold. The results from the Table III shows our proposed model achieved the best performance in all instances for all the classes using the ACC, SE, PR, SP, and AUC evaluation metrics. It can be observed from Table III that SFFT-CapsNet obtained OA, OS, and OP results of 99.0%, 100%, and 99.8%, respectively. This means the results from Table III concludes that the proposed model outperformed all the other state-of-the-art works compared. The second-best emanated from another work presented by Rajagopalan et al., [57] using CNN which obtained 97.0%, and 93.4%, on OA and OS. Though in their paper, the study failed to report the result for OP. The third best model with accuracies of 90.1%, 86.8% and 86.3% for OA, OS, and OP, respectively was also presented in a study known as the Lesion Attention Convolutional Neural Network (LACNN) which was proposed by Leyuan et. al. [53]. In all, the HOG-SVM model achieved the least performance with the accuracies of 78.1%, 65.3%, 460 and 71.8% on OA, OS, and OP, respectively.

Fig. 11 shows Histogram representing of the overall accuracies of OA, OS, and OP for HOG-SVM, Transfer Learning, VGG16, LACNN, IFCNN, LGCNN, CNN by Rajagopalan, and SFFT-CapsNet. Fig. 11 indicates that the

proposed model outperformed the current-state-of-art models in all the metrics tested.

TABLE III. COMPARISON OF RESULTS OF THE SFFT-CAPSNET AND THE STATE-OF-THE-ART WORKS

Method	Classes	AC C (%)	SE (%)	PR (%)	SP (%)	AUC	OA	OS	OP
HOG-SVM [62]	CNV	85.7	87.6	82.0	84.0	92.2			
	DME	91.4	53.8	74.6	97.2	87.3	78.1	65.3	71.8
	DRUSE N	90.2	29.5	52.6	97.0	81.3			
	NORM AL	89.1	90.4	78.1	88.4	94.6			
Transfer Learning [58]	CNV	86.9	76.2	93.9	95.9	96.1			
	DME	91.6	75.5	66.9	94.1	93.8	79.5	7.9	73.1
	DRUSE N	87.2	70.7	42.0	89.1	89.2			
	NORM AL	93.3	88.9	89.5	95.2	98.1			
VGG16 [59]	CVN	91.0	86.6	93.2	94.7	97.2			
	DME	92.8	70.9	74.6	96.2	93.6	83.2	6.2	76.4
	DRUSE N	90.7	54.7	54.5	94.7	88.7			
	NORM AL	91.8	92.6	83.3	91.5	97.2			
LACNN [53]	CNV	92.7	89.8	93.5	95.1	97.7			
	DME	96.6	87.5	86.4	98.0	97.4	90.1	6.8	86.3
	DRUSE N	93.6	72.5	70.0	95.6	93.4			
	NORM AL	<b>97.4</b>	97.3	94.8	97.4	99.2			
IFCNN [60]	CNV	92.4	94.8	87.9	90.9	X			
	DME	94.4	79.2	81.9	97.2	X	87.3	82.5	84.7
	DRUSE N	93.0	64.4	76.8	97.3	X			
	NORM AL	98.4	91.5	92.2	96.4	X			
LGCNN [61]	CNV	93.3	93.3	91.5	93.3	X			
	DME	93.6	85.7	79.4	96.8	X	88.4	84.6	82.9
	DRUSE N	95.4	71.0	65.2	96.0	X			
	NORM AL	94.6	88.5	95.5	97.9	X			
Rajagopalan et. al., [57]	CNV	X	X	X	X	X			
	DME	X	X	X	X	X			
	DRUSE N	X	X	X	X	X	97.0	93.4	X
	NORM AL	X	X	X	X	X			
SFFT-CapsNet [ours]	CNV	100	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>			
	DME	98.8	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	99.0	100	99.8
	DRUSE N	100	<b>97.2</b>	<b>100</b>	<b>100</b>	<b>100</b>			
	NORM AL	97.1	<b>98.7</b>	<b>0.99</b>	<b>100</b>	<b>100</b>			

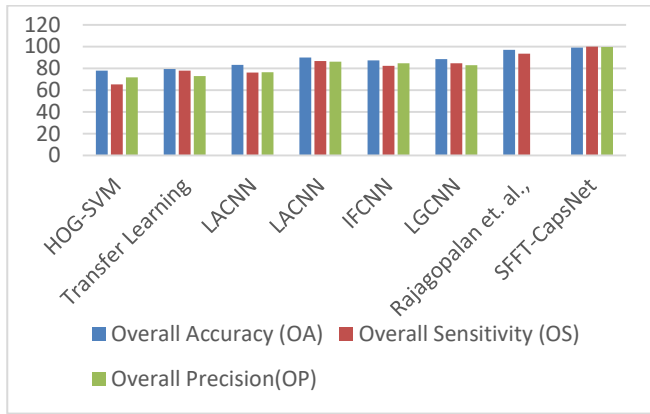


Fig. 11. Histogram representing the overall accuracies of OA, OS, and OP for HOG-SVM, Transfer Learning, VGG16, LACNN, and SFFT-CapsNet.

Fig. 12 shows an evaluation of clusters generated from the raw data and the routing process. Fig. 12(a) represents the raw clustering generated from the input dataset which has no visible clusters. It can be seen that the content is scattered on one cannot observe any possible clusters. Fig. 12(b) also shows the result of the clusters acquired from using the original CapsNet and finally Fig. 12(c) also shows the clusters formed from using the SFFT-CapsNet. From the Fig. 12, it can be visualized that the Fig. 12(c) which was acquired from the proposed model produced better clustering after the routing process than the original CapsNet.

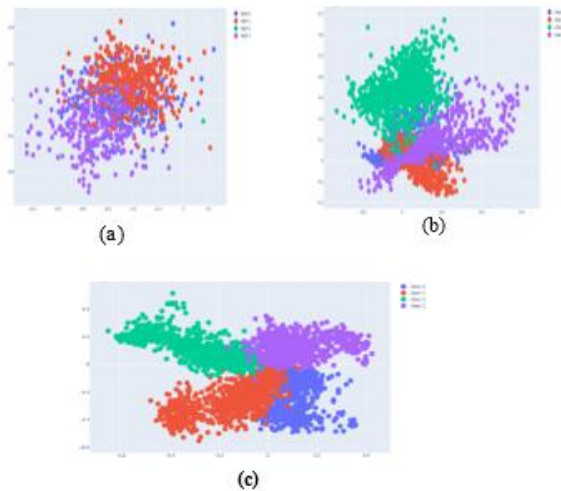


Fig. 12. Visualization of the clusters formed at the retina caps layer (a) raw dataset before routing (b) clusters formed after the routing process of the original CapsNet (c) clusters formed after the routing process of the proposed model.

This can be attributed to the fact that during the routing process, the primary capsules combine with class capsules to establish high agreement which forms better clusters at the retinaCaps layer in the SFFT-CapsNet. The various separations created by the cluster which have been indicated using different colors can be used to determine how efficient the routing process could be and hence determine the efficiency of the model.

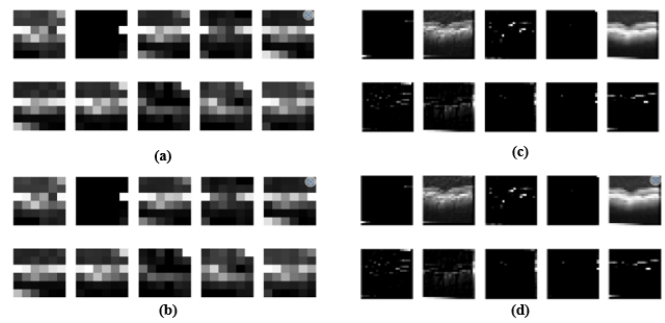


Fig. 13. Activation maps from digit capsule and evaluation capsule of the original CapsNet and SFFT-CapsNet. (a) digit capsule activation from original CapsNet. (b) evaluation capsule activation map from original CapsNet, (c) digit capsule activation from SFFT-CapsNet, and (d) evaluation capsule activation map from SFFT-CapsNet.

Fig. 13 visualizes activation maps performance of the various layers in the original CapsNet and the proposed SFFT-CapsNet as they receive input images. From the visualization, it is observed that the activation from the SFFT-CapsNet provides enough details on how a layer extracts features from the input image to influence the final output of the model. It provides better insight on the actual operations of the CapsNet.

### B. Discussion

The study proposed SFFT-CapsNet for classification of retina OCT images. The result of the model is compared to the original CapsNet which is serving as a baseline for the study. From Table I, the results indicated that the accuracy of proposed SFFT-CapsNet outperformed the baseline model by a difference of 4.8% upon achieving accuracy of 99%. This is a very significant improvement in the field of computer vision. The outstanding performance is due to the novelty employed in the method. Increasing the convolutional layers to six was a good strategy for improving the feature extraction by the model. Secondly, the two enhancement layers introduced in the model improved visibility of hidden patterns and controlled over brightening of the images. Replacing the SoftMax with sigmoid and introducing the power squash prevented the model from generating high activating values that can prevent effective learning of features of the model. From the results of the ablation study in Table II, it can be concluded that such significant improvement was achieved because every layer we introduced strategically had significant impact on the model.

Again, the validation and loss accuracy in Fig. 7 shows that our model achieved higher performance compared to the baseline. Also, Fig. 8(a) and 8(b) show the confusion matrix of the proposed evaluated SFFT-CapsNet models against the original CapsNet. The diagonal outputs indicated in blue illustrated the correct prediction from the models (True positives and true negatives). The misclassifications generated from the model are indicated in white colors as output at the upper and bottom part of the correct predictions. Fig. 8(a) represents the confusion matrix of the SFFT-CapsNet. Fig. 8(b) on the other hand represents the confusion matrix of the original CapsNet. From the confusion matrix, the four instances of the dataset have 242 images each as test samples. The results from the confusion matrix can be concluded that the proposed model obtained the highest correct predictions thus 242, 239, 242, and 235 on CNV, DME, DRUSEN, and

NORMAL, respectively whereas the original CapsNet obtained 231, 230, 239, and 212 on CNV, DME, DRUSEN, and NORMAL, respectively. Fig. 9 presents the ROC-AUC curves and Fig. 10 illustrates the result of Precision-Recall curve. From the results, it can be deduced that the proposed SFFT-CapsNet controls misclassification better than the original CapsNet. From the output of the confusion matrix, our proposed model recorded the least misclassifications when compared to that of the baseline CapsNet. The results of the visualization in Fig. 12 are strong indication that our model is able to generalize well by learning the required features for better classification performance.

The performance of the proposed model can be associated to fact that the introduction of the two layers thus the CLAHE and Fourier transform could sufficiently reduce the noise in the input images. Also, increasing the convolution layer also enhanced the chances of the model extracting required features that could impact on the results of the model.

However, the high misclassification of the original CapsNet can be attributed to the insufficient convolutional layers to extract the required features to support the prediction and classification task. This is in line with the findings proposed by Cao et al., [63] which concluded that CapsNet is unable to perform well on complex images because the convolutional layer is not sufficient able to extract the required features which ends up including features that may not be required and can lead to misclassification.

Also, the performance of our model could compete and outperform the state-of-the-art models that have been implemented on the retina OCT dataset as indicated in Table III. The results indicate a strong confirmation that the proposed SFFT-CapsNet achieved the best performance in all scenarios of the evaluation Matrix. The performance of the CapsNet was not surprising as compared to the other methods which were implemented using CNN. This is because the CapsNet with the dynamic routing algorithm was able to recognize the pose, texture and spatial relationship which contributed to the performance of the classification model.

## V. CONCLUSION

The study proposed an efficient CapsNet architecture for classifying retina OCT images. The study reconstructed the original CapsNet to include two extra layer components thus the contrast limited adaptive histogram equalization layer and Fourier transform layer. The two layers are all image enhancement layers which presented the model with more visibility of the input images. Again, the study increased the convolutional layers to six to ensure better feature extraction and replaced the SoftMax activation function with sigmoid.

Four-class (CNV, DME, DRUSEN, and NORMAL) retina OCT image dataset presented by UCSD were used for training and testing the proposed capsule framework. Evaluations of models were conducted using evaluation metrics such ACC, SE, PR, SP, AUC on the individual Class while OA, OS, and OP to measure the overall performance of the models. The results of the proposed model were compared with that of the original CapsNet in terms of performance accuracies. A further comparative analysis was conducted using the results of the

propose model and that of the state-of-the-arts deep learning standard models that have been applied to the retina OCT image dataset.

The summary of the results has been presented in Table III. The results indicated that the proposed SFFT-CapsNet obtained OA, OS, and OP results of 99.0%, 100%, and 99.8%, respectively which were the best results in all instances compared with the other existing works. This performance indicates that the proposed technique is better in detecting eye diseases from retina OCT images. The method can be adopted to help ophthalmologists in detecting eye disease from retina OCT images. Although the proposed SFFT-CapsNet model achieved high performance compared to the state-of-the-art models, however, it was found that the model still needs improvement. As part of the future works, the study aims to propose an effective activation function that can help the convolutional layers implemented to extract better features required for the model performance. Also, the study seeks to test the final version of the model on complex images to establish the robustness of the model.

## ACKNOWLEDGMENT

The authors would like to thank the editor and the reviewers for their helpful suggestions and valuable comments.

## REFERENCES

- [1] G. Litjens, T. Kooi et al., "A survey on deep learning in medical image analysis". CoRR abs/1702.05747 (2017). <http://arxiv.org/abs/1702.05747>.
- [2] P. Afshar P., A. Oikonomou F. Naderkhani, N. P. Tyrrell, N. Konstantinos Plataniotis, K Farahani and A. Mohammadi, "3D-MCN: A 3D Multi-scale Capsule Network for Lung Nodule Malignancy Prediction" Scientific Reports (2020) 10:7948 | <https://doi.org/10.1038/s41598-020-64824-5>.
- [3] S. Shen., S. X. Han., D. R. Aberle, A. A. Bui and W. Hsu, "An interpretable deep hierarchical semantic convolutional neural network for lung nodule malignancy classification", Expert Systems With Applications 128 (2019) 84–9. <https://doi.org/10.1016/j.eswa.2019.01.048> 0957-4174/ 2019.
- [4] , M. Avendi, A. Kheradvar and H. Jafarkhani, "A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI", Med. Image Anal. 30, 2016. 108–119.
- [5] A. P. Sunija, S. Kar, S. Gayathri, P. Varun Gopi and P. Palanisamy, OctNET: A Lightweight CNN for Retinal Disease Classification from Optical Coherence Tomography Images, Computer Methods and Programs in Biomedicine (2020), doi: <https://doi.org/10.1016/j.cmpb.2020.105877>.
- [6] O. J. P Chary and González OFA., "A Systematic Review of Deep Learning Methods Applied to Ocular Images", Cien.Ing.Neogranadina, vol. 30, no. 1, pp. 9-26, Nov. 2019.
- [7] N. Gurudath, M. Celenk, and H. B. Riley, "Machine Learning Identification of Diabetic Retinopathy from Fundus Images," In 2014 IEEE Signal Processing in Medicine and Biology Symposium (SPMB), 2014, pp. 1-7. doi:10.1109/SPMB.2014.7002949 [ Links ].
- [8] R Priyadarshini, N. Dash, and R. Mishra, "A Novel Approach to Predict Diabetes Mellitus Using Modified Extreme Learning Machine," In 2014 International Conference on Electronics and Communication Systems (ICECS), 2014, pp. 1-5. doi:10.1109/ECS.2014.6892740 [ Links ].
- [9] P Afshar, A Mohammadi, K N Plataniotis, "Brain Tumor Type Classification via Capsule Networks", [C]// 2018 25th IEEE International Conference on Image Processing (ICIP). IEEE, 2018.

- [10] R. A. Welikalaa et al., "Automated Detection of Proliferative Diabetic Retinopathy Using a Modified Line Operator and Dual Classification," *Computer Methods and Programs in Biomedicine*, vol. 114, no. 3, pp. 247-261, 2014. doi:10.1016/j.cmpb.2014.02.010 [ Links ].
- [11] R. LaLonde, "U Bagci. Capsules for Object Segmentation", arXiv:1804.04241, 2018.
- [12] O. Zhicheng Jia, and P. Tae-Eui Kam, "Dynamic Routing Capsule Networks for Mild Cognitive Impairment Diagnosis" 2019. DOI: 10.1007/978-3-030-32251-9\_68.
- [13] M. Mehdy, P. Ng, E. F. Shair, N. Saleh, and C. Gomes, "Artificial Neural Networks in Image Processing for Early Detection of Breast Cancer", *Comput. Math. Methods Med.* 2017, 2610628.
- [14] P. M. Kwabena, A. Felix Adekoya, A. Abra Mighty et al., "Capsule Networks – A survey", *Journal of King Saud University – Computer and Information Sciences*, 2019, <https://doi.org/10.1016/j.jksuci.2019.09.01>.
- [15] J. Rathod, V. Waghmode, A. Sodha and P. Bhavathankar, "Diagnosis of skin diseases using Convolutional Neural Networks" In Proceedings of the 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 29–31 March 2018; pp. 1048–1051.
- [16] J. Naranjo-Torres, M. Mora, R. Hernández-García, R. J. Barrientos, C. Fedes, A. Valenzuela, "A Review of Convolutional Neural Network Applied to Fruit Image Processing", *Appl. Sci.* 2020, 10, 3443.
- [17] N. Noord and E. Postma, "Learning scale-variant and scale-invariant features for deep image classification. *Pattern Recognit*", 61, 583–592, 2017,.
- [18] O. Ronneberger, P. Fischer and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation". arXiv preprint arXiv:1505.04597, 2015.
- [19] J. Gu, "Interpretable Graph Capsule Networks for Object Recognition. The Thirty-Fifth AAAI Conference on Artificial Intelligence", Association for the Advancement of Artificial Intelligence (AAAI-21), 2021.
- [20] E. Xi, S. Bing, Y. Jin, "Capsule Network Performance on Complex Data" (2017).
- [21] G. E. Hinton, A. Krizhevsky and S. D. Wang, "Transforming Auto-Encoders", In *Artificial Neural Networks and Machine Learning—ICANN 2011* Eds.; Springer: Berlin/Heidelberg, Germany, 2011; Volume 6791, pp. 44–51. ISBN 978-3-642-21734-0. [https://doi.org/10.1007/978-3-642-21735-7\\_6](https://doi.org/10.1007/978-3-642-21735-7_6).
- [22] S. Sabour, N. Frosst, and G. E. Hinton. "Dynamic Routing Between Capsules." 2017. arXiv preprint arXiv:1710.09829.
- [23] B. Jia, and Q. Huang, "DE-CapsNet: A diverse enhanced capsule network with disperse dynamic routing", *Appl. Sci.* 10 (884), pp 1–13. 2020.
- [24] D. Wang, A. Khosla, R. Gargeya, H. Irshad, A. H. Beck, "Deep learning for identifying metastatic breast cancer" 2016, arXiv preprint arXiv:1606.05718.
- [25] A. Nithya, A. Appathurai, N. Venkatadri, D. R. Ramji and C. A. Palagan, "Kidney disease detection and segmentation using artificial neural network and multi-kernel k-means clustering for ultrasound images". *Measurement* 149:106952, 2020.
- [26] N. Arunkumar, M. A. Mohammed, S. A. Mostafa, D. A. Ibrahim, J. Rodrigues, V. H. C. de Albuquerque, "Fully automatic model-based segmentation and classification approach for MRI brain tumor using artificial neural networks", *Concurr Comput Pract Exp* 32(1):4962, 2020.
- [27] S. P. K., Karri, D. Chakraborty and J. Chatterjee, "Transfer learning-based classification of optical coherence tomography images with diabetic macular edema and dry age-related macular degeneration" *Biomed. Opt. Express* 8, 579–592, 2017. doi: 10.1364/BOE.8.000579.
- [28] H. Ren, J. Su and H. Lu, "Evaluating generalization ability of convolutional neural networks and capsule networks for image classification via top-2 classification" 2019. ArXiv:1901.10112v2 Cs.CV.
- [29] Z. Yang and X. Wang, "Reducing the dilution: An analysis of the information sensitiveness of capsule network with a practical solution". arXiv, 2019, arXiv:1903.10588.
- [30] J. Liu, F. Gao, R. Lu, Y. Lian, D. Wang, X. Luo, and C., "Wang-DDRM-CapsNet: Capsule Network Based on Deep Dynamic Routing Mechanism for Complex Data". ICANN 2019, LNCS 11727, pp. 178–189, 2019. [https://doi.org/10.1007/978-3-030-30487-4\\_15](https://doi.org/10.1007/978-3-030-30487-4_15).
- [31] Z. Zhao, A. Kleinhans, G. Sandhu, I. Patel and K. P. Unnikrishnan., "Capsule networks with max-min normalization", ArXiv:1903.09662v1 [Cs.CV], 1–15. 2019.
- [32] P. K. Mensah, B. W. Asubam and A. A. Mighty, "Exploring the performance of LBP-capsule networks with KMeans routing on complex images", *Journal of King Saud University – Computer and Information Sciences*, 2020. <https://doi.org/10.1016/j.jksuci.2020.10.006>.
- [33] E. Xi, S. Bing, and Y. Jin. "Capsule network performance on complex" data. 2017. arXiv preprint arXiv:1712.03480.
- [34] Z. Zhang, S. Ye, P. Liao, Y. Liu, G. Su and Y. Sun, "Enhanced Capsule Network for Medical image classification", 978-1-7281-1990-8/20/\$31.00 IEEE, 2020.
- [35] F. Shaikat, G. Raja, R. Ashraf, S. Khalid, M. Ahmad and A. Ali, "Artificial neural network based classification of lung nodules in CT images using intensity, shape and texture features". *J Ambient Intell Hum Comput* 10: pp 4135–4149, 2019.
- [36] H. P. Nguyen and B. Ribeiro, "Advanced Capsule Networks via Context Awareness", ICANN 2019, LNCS 11727, pp. 166–177, 2019. [https://doi.org/10.1007/978-3-030-30487-4\\_14](https://doi.org/10.1007/978-3-030-30487-4_14).
- [37] C. Xiang, L. Zhang, Y. Tang, W. Zou and X. Chen, "MS-CapsNet: a novel multi-scale capsule network", *IEEE Signal Process Lett* 25(12):1850–1854, 2018.
- [38] S. S. R. Phaye, A. Sikka, A. Dhall and D. Bathula, "Dense and diverse capsule networks: Making the capsules learn better", ArXiv : 1805 . 04001v1 [ Cs . CV ] pp 1–11. 10 May 2018.
- [39] 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, CVPR, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
- [40] G. Larsson, M. Maire, and G. Shakhnarovich, "FractalNet: Ultra-deep neural networks without residuals", ArXiv:1605.07648v4 [ Cs.CV], pp1–11. 2017. Retrieved from <http://arxiv.org/abs/1605.07648>.
- [41] G. Deborshi and R. Sun, "Application of capsule networks for image classification on complex datasets". 2019.
- [42] S. B. S Bhamidi and M. El-Sharkawy, "Residual capsule network. 2019 IEEE 10th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference", UEMCON 2019, 0557–0560. <https://doi.org/10.1109/UEMCON47517.2019.8993019>.
- [43] Y. Li, M. Qian, P. Liu, Q. Cai, X. Li X, J. Guo, H. Yan et al "T recognition of rice images by UAV based on capsule network". *Clust Comput* 22(4):9515–9524, 2019.
- [44] M. E. Paoletti, J. M. Haut, R. Fernandez-Beltran, J. Plaza, A. Plaza and J. Li and F. Pla, "Capsule networks for hyperspectral image classification", *IEEE Trans Geosci Remote Sens* 57(4):2145–2160. 2018.
- [45] F. Deng, P. Shengliang, X. Chen, Y. Shi, T. Yuan and P. Shengyan, "Hyperspectral image classification with capsule network using limited training samples". *Sensors* 18(9):3153, 2018.
- [46] W. Y. Wang, H. C. Li, L. Pan, G. Yang and Q. Du, "Hyperspectral image classification based on capsule network". In: *IGARSS 2018 IEEE international geoscience and remote sensing symposium*. IEEE, pp 3571–3574, 2018.
- [47] K. Adu, Y. Yu, J. Cai and N. Tashi, "Dilated Capsule Network for Brain Tumor Type Classification Via MRI Segmented Tumor Region". 2019 IEEE International Conference on Robotics and Biomimetics, 2020. DOI: 10.1109/ROBIO49542.2019.8961610.
- [48] H. J. D. Koresh and S. "Chacko. Classification of noiseless corneal image using capsule networks. *Soft Computing*, 2020" <https://doi.org/10.1007/s00500-020-04933-5>.

- [49] J. Wang, G. Deng, W. Li, C. Yiwei, G. Feng, H. Liu, Y. He, G. Shi, "Deep learning for quality assessment of retinal OCT images", *Biomedical Optics Express*, 2019, <https://doi.org/10.1364/BOE.10.006057>
- [50] P. A. Keane, P. J. Patel, S. Liakopoulos, F. M. Heussen S. R, Sadda, A. Tufail. "Evaluation of age-related macular degeneration with optical coherence tomography" *Surv Ophthalmol.* 2012; 57: 389e414. 5.
- [51] T. Ilginis, J. Clarke and P. J. Patel. "Ophthalmic imaging". *Br Med Bull.* 2014;111(1):77-88. doi:10.1093/bmb/ldu022.
- [52] C. Neely, K. J. Bray, C. E. Huisingh, M. Clark, G. J. McGwin, and C. Owsley, "Prevalence of undiagnosed age-related macular degeneration in primary eye care," *JAMA Ophthalmol.*, vol. 135, no. 6, pp. 570-575, 2017
- [53] F. Leyuan, C. Wang, S. Li, H. Rabbani, X. Chen, and Z Liu, "Attention to Lesion: Lesion-Aware Convolutional Neural Network for Retinal Optical Coherence Tomography Image Classification", 2019. DOI 10.1109/TMI.2019.2898414
- [54] T. Hahn, M. Pyeon and G Kim, "Self-routing capsule networks". In: Wallach HM, Larochelle H, Beygelzimer A, d'Alch'e-Buc F, Fox EB, Garnett R, editors. *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, 8-14 December 2019, Vancouver, BC, Canada.* 2019. p. 7656-65. URL: <http://papers.nips.cc/paper/8982-self-routing-capsule-networks>.
- [55] Malmgren C. A, "Comparative Study of Routing Methods in Capsule Networks. Master's thesis" Linköping University, Computer Vision; 2019.
- [56] J. E. Lenssen, M. Fey, P. Libuschewski, "Group equivariant capsule networks", In: Bengio S, Wallach H, Larochelle H, Grauman K, Cesa-Bianchi N, Garnet R, editors. *Advances in Neural Information Processing Systems 31.* Curran Associates, Inc.; 2018. p. 8844-53. URL: <http://papers.nips.cc/paper/8100-group-equivariant-capsule-networks.pdf>.
- [57] N. N. V. Rajagopalan, and A. N. Josephraj, "Diagnosis of retinal disorders from optical coherence tomography images using cnn", *PLOS ONE* 16 (7) (2021),1-17.doi:10.1371/journal.pone.0254180.
- [58] D. S. Kermany et al., "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122-1131, 2018.
- [59] K. Simonyan, A. "Zisserman, Very deep convolutional networks for large-scale image recognition", In: *ICLR 2015 Conference Proceedings*, pp. 1-14. ArXiv:1409.1556v6 [Cs.CV] 2015.
- [60] F. Leyuan, Y. Jin, L. Huang, S. Guo, G. Zhao and X. Chen, "Iterative fusion convolutional neural networks for classification of optical coherence tomography images," *J. Vis. Commun. Image Represent.*, vol. 59, pp. 327- 333, 2019.
- [61] L. Huang,, X. He,, L. Fang, H. Rabbani, and X. Chen, "Automatic classification of retinal optical coherence tomography images with layer guided convolutional neural network". *IEEE Signal Process. Lett.* 26, 1026-1030. doi: 10.1109/LSP.2019.2917779
- [62] P. P. Srinivasan et al ., "Fully Automated Detection of Diabetic Macular Edema and Dry Age-Related Macular Degeneration from Optical Coherence Tomography Images," *Biomedical Optics Express*, vol. 5, no. 10, pp. 3568-3577, 2014. doi:10.1364/BOE.5.003568
- [63] S. Cao, Y. Yao, G An, E2-capsule neural networks for facial expression recognition using AU-aware. *IET Image Process. Electron. Lett.*, 1-2 2019. <https://doi.org/10.1049/iet-ipr.2020.0063>



# Deep Neural Network-based Detection of Road Traffic Objects from Drone-Captured Imagery Focusing on Road Regions

Hoanh Nguyen

Faculty of Electrical Engineering Technology, Industrial University of Ho Chi Minh City, Ho Chi Minh City, Vietnam

**Abstract**—This paper presents a novel deep learning approach for the detection of traffic objects from drone-based imagery, focusing predominantly on the rapid and accurate detection of vehicles within road sections. The proposed method consists of two primary components: a road segmentation module and a vehicle detection network. The former leverages a residual unit with skip-connections to effectively extract road areas, while the latter employs a modified version of the YOLOv3 architecture, tailored for high-accuracy and high-speed vehicle detection. To address the issue of data imbalance, which is a pervasive challenge in drone images, this paper utilizes a range of data augmentation techniques to improve the robustness of the proposed model. Experimental results on the UAVDT and UAVid datasets exhibit that the proposed model attains a substantial boost in accuracy and inference speed of vehicle detection in comparison to the existing methods. These findings underscore the potential of the proposed approach for real-world traffic monitoring applications, where rapid and reliable vehicle detection is paramount.

**Keywords**—Deep learning; drone images; vehicle detection; road segmentation; data imbalance

## I. INTRODUCTION

As drone technology has rapidly advanced in recent years, numerous practical applications based on images collected from drones have been developed. Among these, the most notable are intelligent processing applications based on images obtained from drones, such as object detection [1-2], object segmentation [3], traffic analysis [4], traffic prediction [5], and work monitoring systems [6]. Compared to ground-collected images, drone-collected images often have many more advantages, such as encompassing information from a vast area, dynamic coverage, and different altitudes and positions. Due to these benefits, processing based on drone images often faces many challenges. These challenges stem from various factors including complex backgrounds, a global perspective, and varying scales of targets. Fig. 1 describes some cases of drone images that pose many challenges for object detection and segmentation tasks. More specifically, objects in drone images are often obscured or overlap with other objects. The number of objects in drone images is usually very large, and the size of the objects in the images is often small.

Given the significant advancements in deep learning, particularly in convolutional neural networks (CNN), numerous methods have been introduced in recent years to address object detection and segmentation using drone images

and CNN. In [7], the authors propose an automatic image annotation method, analyze YOLOv3's training behavior on the natural UAVDT dataset, and demonstrate the performance that can be achieved through synthetic training, as well as how synthetic augmentation can enhance the natural training set's performance. Kyrkou et al. [8] present a comprehensive approach to developing a single-shot object detector based on CNN for UAVs, specifically focusing on vehicle detection in resource-constrained environments. The paper covers the entire development process including data collection, training, CNN architecture design, and optimizations for efficient deployment on lightweight embedded processing platforms suitable for drone images. Li et al. [9] introduced the Density-Map guided object detection Network (DMNet) as an inventive approach to tackle the complexities of object detection in high-resolution aerial photos, particularly issues related to vast differences in object size and irregular object distribution. The DMNet, which integrates a density map generation module, an image cropping module, and an object detector, uses pixel intensity to establish a subtle boundary for image cropping and to discern object scales. In [10], the authors propose a separate resampling algorithm to alter the input test images' size and subsequently extend the object's impact in deeper layers of the detection model. They utilize a pre-trained Faster R-CNN [11] object detection model with Inception-V2 [12], applying transfer learning to submeter satellite images with passenger vehicles as the target objects. In [13], the author present a novel vision-based system for vehicle detection and counting on highways, aiming to address the challenge of detecting vehicles of varying sizes. This system utilizes a novel segmentation technique to partition the highway road surface in the image into distant and near areas. Subsequently, it leverages the YOLOv3 network for vehicle detection. Recently, Feng et al. [14] suggested utilizing the mean classification score as a metric for gauging the classification accuracy of each category during training. They introduced the Equilibrium Loss (EBL) and Memory-augmented Feature Sampling (MFS) techniques to ensure balanced classification. Together, EBL and MFS notably enhance detection performance for less represented classes while either preserving or boosting performance for the more prevalent ones. In addition to the methods mentioned above, references [15-18] provide systematic reviews of object detection and segmentation methods based on drone images.

While the aforementioned methods have achieved certain successes, there remain numerous issues that need to be

addressed to construct an effective model for object detection based on drone images. This paper introduces a proficient model for detecting objects in drone images. Aiming to provide efficient data for intelligent traffic monitoring applications, the model proposed in this paper performs vehicle object detection based on regions of interests (RoIs) rather than on the entire image. Detecting objects based on RoIs helps to eliminate unrelated areas during processing, not only increasing the model's accuracy but also significantly enhancing execution speed, particularly for high-resolution images obtained from drones. To efficiently create RoIs, specifically road sections, this paper proposes applying a segmentation model for road detection. Based on extracted road sections, a method is proposed to enhance both the accuracy and inference speed of vehicle detection from road sections. Moreover, in response to the widespread issue of data imbalance often encountered in drone imagery, this paper applies an assortment of data augmentation strategies aimed at enhancing the resilience and reliability of the proposed model. Finally, this paper also proposes using suitable models and datasets that meet the requirements.

The paper is organized as follows: Section II delves into the details of the proposed methodology. Section III presents the results derived from the framework's implementation. Finally, Section IV concludes the paper and highlights potential avenues for future research.



Fig. 1. Some images illustrate the challenges of object detection and segmentation tasks with images from drones.

## II. METHODOLOGY

### A. Overview of the Proposed Method

Fig. 2 illustrates the overall structure of the model proposed for the problem of road traffic object detection from drone images. Aiming at detecting objects on the road, specifically vehicles, with high inference speed to provide real-time information for road management systems, a deep learning network is first used for road detection and segmentation. Extracting the road sections helps the model focus on detecting objects on the road, thereby not only significantly increasing the inference speed of the overall model but also improving accuracy. The input to this deep learning network is the input images, and the output is the predicted road sections. Based on the extracted road sections, a deep learning network based on the YOLOv3 architecture is designed for object detection, specifically vehicles on the road. Using a vehicle detection model based on the YOLOv3 architecture significantly

improves the model's inference time, especially with images obtained from drones where the number of objects on the road is considerable. Additionally, the paper also proposes using data augmentation strategies to address issues related to data imbalance often encountered in drone imagery. Details about each proposed network will be presented in the following sections.

### B. Road Segmentation

This paper approaches road segmentation in images as a binary segmentation task, classifying each pixel in the input image as either part of the road or the background. Several models have been proposed for segmentation, such as UNet [19], Segnet [20], DeepLabv3+ [21], or the more recent DoubleUNet [22]. These models typically combine an encoder for feature extraction with a decoder for segmentation. The encoder is critical in extracting features, capturing contextual information, reducing dimensionality, creating a hierarchical representation, and making use of transfer learning. Conversely, the decoder is responsible for upsampling, reconstructing the feature maps, refining the segmentation output with contextual information, applying non-linear mappings, and generating the final segmentation output through multi-level feature fusion and skip connections. In pursuit of high accuracy and fast inference speed, this paper proposes the use of the ResNet50 model [23] as the encoder and a combination of residual blocks and other operations [24] as the decoder, as depicted in Fig. 3. Specifically, this paper utilizes the ResNet50 model, which is pre-trained on the ImageNet dataset [25], to ensure smoother convergence and elevate the overall performance. The decoder consists of four blocks, each including upsampling, concatenation, and skip-connections operations. In detail, each block's input feature map is initially upsampled to a higher resolution using transpose convolution. Following this, a concatenation operation merges the upsampled feature map with its encoder counterpart. A residual unit with skip-connections then produces the final feature map for that block. Opting for residual blocks over standard convolutional ones streamlines network training and guarantees undegraded information propagation due to the skip connections. The decoder's final block output undergoes a  $1 \times 1$  convolution layer and a sigmoid function, resulting in the output segmentation map.

For training the road segmentation network, this paper employs Dice Loss [26] as the main loss function. Dice Loss is particularly useful in segmentation tasks where the classes are imbalanced. The goal of the road segmentation task is to categorize each image pixel as either road or background. Given that the number of pixels associated with roads is typically far less than those tied to the background, this presents a significantly imbalanced problem. Dice Loss helps mitigate this issue by maintaining a balance between the foreground and the background. Mathematically, the Dice Loss can be defined as:

$$L_d = 1 - 2 \frac{y_{pre} \cdot y_{gt} + \epsilon}{y_{pre} + y_{gt} + \epsilon} \quad (1)$$

where  $y_{pre}$  and  $y_{gt}$  represent the predicted and ground truth, respectively.  $\epsilon$  is a minute positive quantity employed to prevent a division by zero issue.

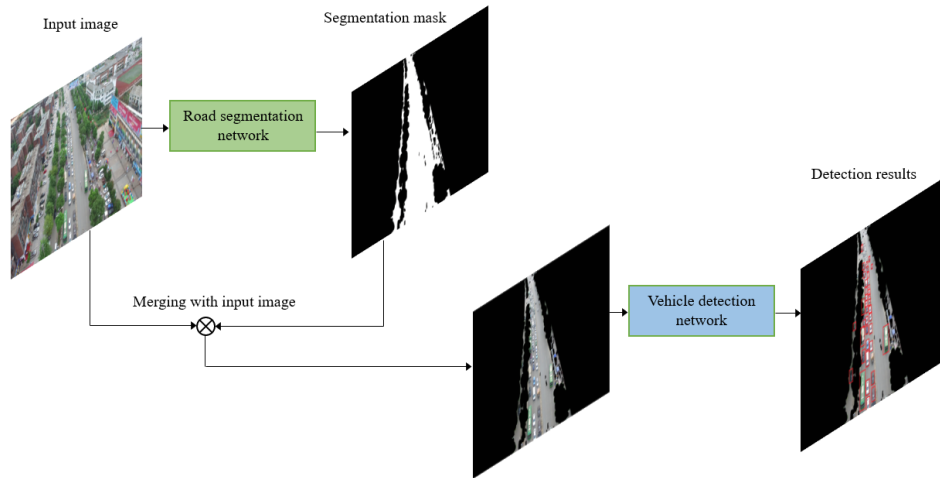


Fig. 2. The structure of the proposed approach.

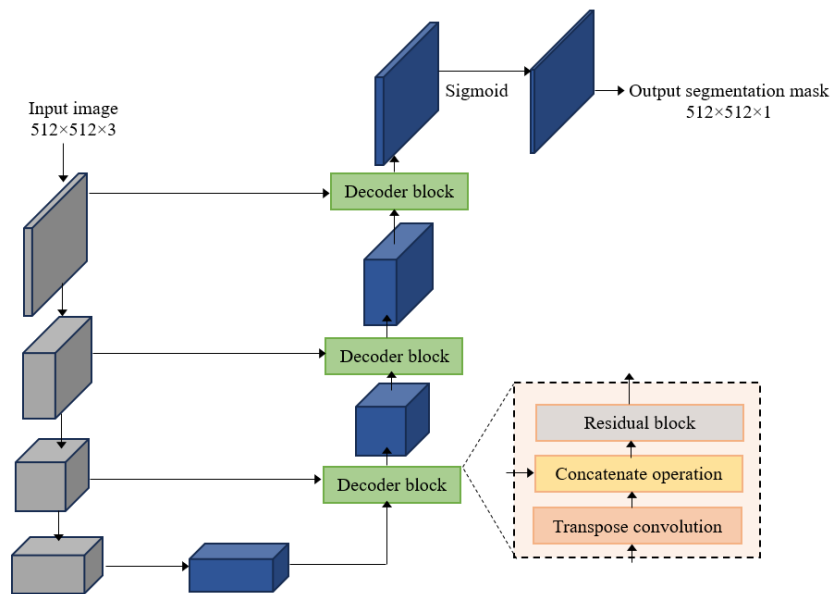


Fig. 3. Encoder-decoder structure for road segmentation.

### C. Vehicle Detection based on Road Sections

Aiming for quick and accurate vehicle detection based on road sections extracted from input images, especially for small vehicles, this paper proposes a vehicle detection model based on the YOLOv3-tiny architecture [27]. Specifically, modifications were made to the YOLOv3-tiny model based on two criteria: model size and its capability to detect small objects. In CNN architecture, deeper layers containing larger numbers of channels and smaller sizes typically store rich semantic information, beneficial for object classification. Conversely, shallower layers with fewer channels and larger sizes typically house rich spatial information, useful for preserving object structure details. Since vehicle detection task only distinguishes between vehicles and background classes, it is significantly simpler than generic object detection. As a result, the shallower network layers can be reduced in the number of channels to decrease the model's complexity without

impacting its accuracy. Based on these analyses, this paper implemented changes to the structure of the first convolutional layers of the YOLOv3-tiny architecture. Specifically, the number of filters in the first two convolutional layers was reduced to three. The rest of the convolutional layers maintained their original number of filters. Table I details the structure of the proposed model in this paper and the original YOLOv3-tiny model. The detection head makes predictions on two feature maps with scales of  $13 \times 13$  and  $26 \times 26$ . With these modifications, the proposed model can significantly reduce computation costs while maintaining network performance. Additionally, the filter size in the detection layers was changed from  $3 \times 3$  to  $1 \times 1$ , improving the model's nonlinearity and aiding the detection model in learning difficult samples. With these changes, the new model has reduced FLOPs to 2.5BFLOPs compared to the original model's 5.4BFLOPs, while the model size has shrunk to 20MB compared to the original 34MB.

TABLE I. COMPARING THE STRUCTURE OF THE ORIGINAL YOLOV3-TINY MODEL AND THE MODEL PROPOSED IN THE PAPER

Layer	Original YOLOv3-tiny			Proposed architecture		
	Type	Filter	Output	Type	Filter	Output
0	Convolutional	3×3×16	416×416×16	Convolutional	3×3×3	416×416×3
1	Max Pooling	2×2	208×208×16	Max Pooling	2×2	208×208×3
2	Convolutional	3×3×32	208×208×32	Convolutional	3×3×3	208×208×3
3	Max Pooling	2×2	104×104×32	Max Pooling	2×2	104×104×3
4	Convolutional	3×3×64	104×104×64	Convolutional	3×3×64	104×104×64
5	Max Pooling	2×2	52×52×64	Max Pooling	2×2	52×52×64
6	Convolutional	3×3×128	52×52×128	Convolutional	3×3×128	52×52×128
7	Max Pooling	2×2	26×26×128	Max Pooling	2×2	26×26×128
8	Convolutional	3×3×256	26×26×256	Convolutional	3×3×256	26×26×256
9	Max Pooling	2×2	13×13×256	Detection		
10	Convolutional	3×3×512	13×13×512	Max Pooling	2×2	13×13×256
11	Convolutional	1×1×256	13×13×256	Convolutional	3×3×512	13×13×512
12	Convolutional	3×3×255	13×13×255	Detection		
13	Detection					

D. Data Augmentation Strategy

Since data imbalance among classes presents a significant challenge for vision tasks based on drone images, this paper proposes several data augmentation techniques to address this issue. Fig. 4 displays the outcomes of the data augmentation techniques applied in this paper on a consistent input image. The strategies employed to augment drone image data in this study encompass random erasing, random rotation, random brightness, random cropping, and random zoom.

1) *Random erasing*: Random erasing [28] involves selecting a rectangular area within an image at random and replacing its pixels with arbitrary values. This region is determined using a uniform distribution, with both the area and aspect ratios chosen randomly. When parts of the input image are randomly erased during training, it compels the model to develop more adaptable and robust representations. This means the model has to identify the correct class without

depending solely on the complete image, enhancing its focus on pertinent sections of the input. This can enhance the model's generalization capabilities, potentially leading to superior performance on unfamiliar data.

2) *Random rotation*: Random rotation is achieved by rotating the image by a random degree between -90 and +90 degrees. By randomly rotating the image, the model is encouraged to learn to recognize the object in different orientations. This makes the model more robust to the orientation of objects in the input data. Let  $(x, y); (x', y')$  be the coordinates of the bounding boxes before and after implementing random rotation. In that case,

$$\begin{cases} x' = (x - M_x) \cdot \cos\alpha - (y - M_y) \cdot \sin\alpha + M_x \\ y' = (x - M_x) \cdot \sin\alpha - (y - M_y) \cdot \cos\alpha + M_y \end{cases} \quad (2)$$

where  $\alpha$  is rotation angle and  $(M_x, M_y)$  is the center coordinate of the input image.

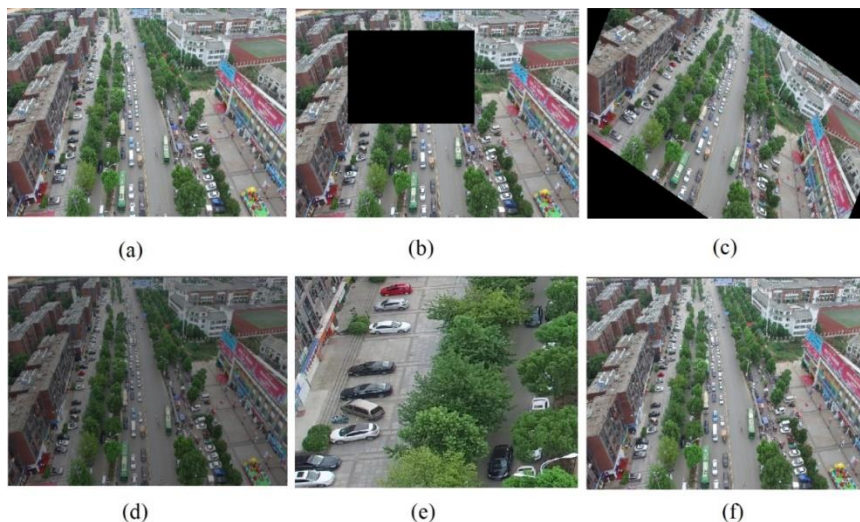


Fig. 4. Data augmentation used in this paper. (a) Original image; (b) Random erasing; (c) Random rotation; (d) Random brightness; (e) Random crop; (f) Random zoom.

3) *Random brightness*: Random brightness involves adjusting the brightness of an image by a random factor. It's usually achieved by converting the image to the HSV color space, adding a random value to the V (Value) channel, and then converting back to the original color space. By modifying the brightness of the image, the model can be trained to be invariant to different lighting conditions. This means that the trained model can recognize an object or feature in an image regardless of whether the image is bright, normally lit, or dim.

4) *Random crop*: Random crop involves selecting a random subsection of the input image for training. The cropped region is smaller than the original image and is resized to the input dimensions of the model. By training the model on a diverse set of cropped images, it can learn to focus on different parts of an object and recognize an object even if only part of it is visible. It can also help to mitigate overfitting as the model cannot rely on the position of features in the image.

5) *Random zoom*: Random zoom involves randomly zooming into or out of an image by a certain amount. This is typically done by resizing the image (upscaling or downscaling) and then cropping or padding it to match the original dimensions. By randomly zooming in or out, the model can learn to recognize objects or features at different scales. It makes the model more scale-invariant, which is beneficial when objects in the test data may appear at different sizes than in the training data. The coordinates of the bounding boxes are updated after implementing random zoom as follows:

$$\begin{cases} x' = \frac{w}{z} + \frac{(x-\frac{w}{2})}{R} \\ y' = \frac{h}{z} + \frac{(y-\frac{h}{2})}{R} \end{cases} \quad (3)$$

where  $R$  is zoom ration and  $(w, h)$  is the width and height of the input image.

### III. RESULTS

#### A. Dataset

This paper utilizes distinct datasets for different tasks as specified in Table II. Specifically, the UAVDT dataset [29] is used to train the vehicle detection network. This dataset is tailored for vehicle detection and tracking tasks, encompassing three categories: car, truck, and bus. For the vehicle detection evaluation in this study, all classes are grouped under a singular category termed 'vehicle'. The images feature a resolution of 1080×540 pixels and capture diverse typical scenes, including squares, arterial roads, and toll stations. For training the road segmentation network, the UAVid dataset [30] is employed. UAVid consists of 300 drone images with resolutions of 4096×2160 or 3840×2160 pixels, captured at a slanted angle, enhancing the intricacy and scale variance of urban street scenes with complex foreground-background elements. Given the substantial image sizes, this paper derives 10,000 random, non-overlapping 512×512 patches from the UAVid dataset. Of these, 8,000 are designated as the training set, while the remaining 2,000 are split equally between the

validation and testing sets. For the purpose of road segmentation, only the annotations relevant to roads are used for training and evaluation in the road segmentation network. Additionally, the UAVid dataset is also used for a joint evaluation of the proposed model. In this evaluation setting, only road annotations are employed throughout the paper, and images lacking road annotations are excluded from the dataset. Furthermore, this study manually uses all vehicle annotations within road sections for object detection training and evaluation.

TABLE II. DATASETS USED IN THIS PAPER

Dataset	Road segmentation	Vehicle detection	Joint segmentation and detection
UAVDT [29]		√	
UAVid	√		√

#### B. Implementation Details

All road segmentation and vehicle detection networks are trained on a NVIDIA RTX 4080 GPU with the support of the PyTorch library. For the road segmentation network, the ResNet50 model, pretrained on the ImageNet dataset, is used as the baseline encoder. This enhances the precision of feature extraction, consequently improving segmentation performance. To boost training performance, an initial learning rate of  $1e^{-4}$  is used to update the parameters, which is then reduced to  $1e^{-7}$  after six consecutive epochs to achieve a better loss rate. The Adam optimizer [31] is utilized to fine-tune the model. The model is trained over 20 epochs with a batch size of 16. For the vehicle detection network, training is carried out using default configurations with a few minor modifications. More specifically, the DarkNet model [27] is deployed, and the SGD optimizer with momentum and weight decay factors of 0.9 and 0.001 respectively is used in the detector training process. The vehicle detection model is trained for 100 epochs with a batch size of 32. A step learning schedule is also employed to gradually reduce the learning rate.

#### C. Road Segmentation Results

This paper conducts experiments with various models to evaluate the effectiveness of the proposed model for the road segmentation task. The compared models are based on EfficientNet [32] and MobileNetv2 [33] as encoders, while the decoders are networks such as DeepLabV3, FPN [34], and Unet. Experiments are performed on the same UAVid dataset with identical training and testing sets. Fig. 5 illustrates the results of the segmentation models used in the experiments, including the model proposed in this paper. It can be seen that the proposed model achieves the best inference speed, while maintaining accuracy comparable to the other models. In terms of accuracy, the DeepLabV3 with EfficientNet model performs the best. However, this model requires 8.2ms as inference time, which is not suitable for intelligent transportation systems requiring real-time processing. Additionally, the results in Fig. 5 also show that models using complex decoder structures, with many layers like DeepLabV3, often require longer processing times due to higher computational demands. The comparison results show that designing an encoder-decoder model that leverages a residual unit with skip-connections, as

proposed in this paper, is very effective both in terms of accuracy and inference speed for the road segmentation task.

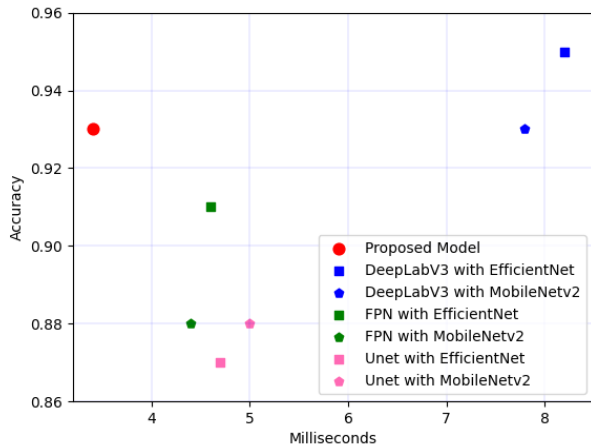


Fig. 5. Performance of different models on the UAVid dataset.

#### D. Vehicle Detection Results

To evaluate the vehicle detection network proposed in this paper, several models have been tested on the UAVDT dataset, including YOLOv3, YOLOv3-tiny, YOLOv4 [35], and SSD-MobileNet [36]. Table III presents the vehicle detection performance of various models on the UAVDT dataset, with the metrics being the Average Precision (AP) in percentage and the speed in milliseconds. From the results, it is clear that YOLOv4 exhibits the highest AP of 86.4, closely followed by YOLOv3 with 82.1. However, when it comes to speed, YOLOv4 falls short with a processing time of 10.4ms, compared to YOLOv3's 12.2ms. On the other hand, YOLOv3-tiny, known for its lightweight architecture, posts a decent AP of 69.4 but shines in speed with a processing time of 6.4ms. The proposed model, a modification of the YOLOv3-tiny architecture, was designed specifically to optimize the model size and improve detection of small objects. Despite achieving a slightly lower AP of 76.2 compared to the original YOLOv3 models, it considerably outperforms all other models in terms of speed with an impressive 4.2ms. This result reflects the efficiency of the proposed design in balancing both precision and speed. In comparison to SSD-MobileNet, which has an AP of 80.4 and speed of 10.6ms, the proposed model excels in inference speed, showing its potential for real-time applications. Therefore, the proposed model offers a promising approach for traffic monitoring, where speed and accurate vehicle detection is crucial.

TABLE III. VEHICLE DETECTION PERFORMANCE OF DIFFERENT MODELS ON THE UAVDT DATASET

Models	AP (%)	Speed (ms)
YOLOv3	82.1	12.2
YOLOv3-tiny	69.4	6.4
YOLOv4	86.4	10.4
SSD-MobileNet	80.4	10.6
Proposed model	76.2	4.2

#### E. Joint Evaluation Results

For joint evaluation, the road segmentation and vehicle detection networks have been integrated to carry out the task of vehicle detection within road sections. Based on the UAVid dataset, labels have been modified to include only vehicles in road sections to determine how accurately the combined model can detect vehicles in these areas. Table IV presents the overall results of several models, including Unet + YOLOv3 and Unet + YOLOv3-tiny. In Table IV, two parts of the comparison are introduced, which include the detection of vehicles in road sections and the detection of vehicles across the entire image. It can be seen that the detection of vehicles in road sections significantly improves the accuracy and inference speed of all models compared to vehicle detection across the entire image. This can be explained by the fact that by focusing only on the necessary parts of the image during detection, the computational cost and the number of objects that need to be predicted are substantially reduced. These findings suggest that intelligent transportation applications could leverage these results to build more efficient systems in their design, thereby facilitating easier system development.

TABLE IV. JOINT EVALUATION RESULTS ON THE UAVID DATASET

Models	AP (%)		Speed (ms)	
	Road sections	Entire image	Road sections	Entire image
Unet + YOLOv3	76.4	62.8	11.6	16.2
Unet + YOLOv3-tiny	62.1	54.4	8.4	10.3
Proposed model	74.1	60.5	6.9	8.4

#### IV. CONCLUSIONS

This paper has designed a novel deep learning method for the detection of traffic objects from drone-based imagery, specifically focusing on the rapid and accurate detection of vehicles within road sections. The proposed method consists of two key components: a road segmentation network and a vehicle detection network. The segmentation network utilizes a residual unit with skip-connections to effectively predict road areas, while the vehicle detection network leverages a modified version of the YOLOv3 architecture, fine-tuned for high-accuracy and high-speed vehicle detection. Moreover, this study addressed the challenge of data imbalance inherent in drone images by implementing various data augmentation techniques, thereby enhancing the model's robustness. The experimental results achieved on the UAVDT and UAVid datasets highlighted the effectiveness of the proposed model. It not only enhanced the accuracy of vehicle detection but also improved the inference speed as compared to existing methods. These results highlight the potential of the proposed approach for practical traffic monitoring applications, where rapid and accurate vehicle detection is of utmost importance. However, it's important to note that the proposed model's effectiveness may be limited to adverse weather conditions or low-light scenarios, as it heavily relies on visual data captured by drones, which can be adversely affected by such factors. Additionally, the model's performance might degrade when applied to highly congested traffic scenes with overlapping vehicles, posing challenges in accurate object detection. For future work, this

paper plans to extend this approach to the detection of more diverse traffic objects beyond vehicles, such as pedestrians and cyclists.

## REFERENCES

- [1] Patrik, Aurello, Gaudy Utama, Alexander Agung Santoso Gunawan, Andry Chowanda, Jarot Sembodo Suroso, and Widodo Budiharto. "Modeling and implementation of object detection and navigation system for quadcopter drone." *ICIC Express Letters* 13, no. 6 (2019): 461-468.
- [2] Arrahmah, Annisa Istiqomah, Rissa Rahmania, and Dany Eka Saputra. "Comparison between convolutional neural network and K-nearest neighbours object detection for autonomous drone." *Bulletin of Electrical Engineering and Informatics* 11, no. 4 (2022): 2303-2312.
- [3] Eerapu, Karuna Kumari, Shyam Lal, and A. V. Narasimhadhan. "O-SegNet: Robust encoder and decoder architecture for objects segmentation from aerial imagery data." *IEEE Transactions on Emerging Topics in Computational Intelligence* 6, no. 3 (2021): 556-567.
- [4] Liu, Shuai, Xin Li, Huchuan Lu, and You He. "Multi-object tracking meets moving UAV." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8876-8885. 2022.
- [5] Yin, Xueyan, Genze Wu, Jinze Wei, Yanming Shen, Heng Qi, and Baocai Yin. "Deep learning on traffic prediction: Methods, analysis, and future directions." *IEEE Transactions on Intelligent Transportation Systems* 23, no. 6 (2021): 4927-4943.
- [6] Casierra, Cristian Benjamín García, Carlos Gustavo Calle Sánchez, Javier Ferney Castillo García, and Felipe Muñoz La Rivera. "Methodology for Infrastructure Site Monitoring using Unmanned Aerial Vehicles (UAVs)." *International Journal of Advanced Computer Science and Applications* 13, no. 3 (2022).
- [7] Krump, Michael, Martin Ruß, and Peter Stütz. "Deep learning algorithms for vehicle detection on UAV platforms: first investigations on the effects of synthetic training." In *Modelling and Simulation for Autonomous Systems: 6th International Conference, MESAS 2019, Palermo, Italy, October 29–31, 2019, Revised Selected Papers* 6, pp. 50-70. Springer International Publishing, 2020.
- [8] Kyrkou, Christos, George Plastiras, Theocharis Theocharides, Stylianos I. Venieris, and Christos-Savvas Bouganis. "DroNet: Efficient convolutional neural network detector for real-time UAV applications." In *2018 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, pp. 967-972. IEEE, 2018.
- [9] Li, Changlin, Taojiannan Yang, Sijie Zhu, Chen Chen, and Shanyue Guan. "Density map guided object detection in aerial images." In *proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pp. 190-191. 2020.
- [10] Mansour, Ahmad, Wessam M. Hussein, and Ehab Said. "Small objects detection in satellite images using deep learning." In *2019 Ninth International Conference on Intelligent Computing and Information Systems (ICICIS)*, pp. 86-91. IEEE, 2019.
- [11] Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. "Faster r-cnn: Towards real-time object detection with region proposal networks." *Advances in neural information processing systems* 28 (2015).
- [12] Szegedy, Christian, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. "Rethinking the inception architecture for computer vision." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818-2826. 2016.
- [13] Song, Huansheng, Haoxiang Liang, Huaiyu Li, Zhe Dai, and Xu Yun. "Vision-based vehicle detection and counting system using deep learning in highway scenes." *European Transport Research Review* 11, no. 1 (2019): 1-16.
- [14] Feng, Chengjian, Yujie Zhong, and Weilin Huang. "Exploring classification equilibrium in long-tailed object detection." In *Proceedings of the IEEE/CVF International conference on computer vision*, pp. 3417-3426. 2021.
- [15] Bisio, Igor, Chiara Garibotto, Halar Haleem, Fabio Lavagetto, and Andrea Sciarone. "A systematic review of drone based road traffic monitoring system." *IEEE Access* (2022).
- [16] Chen, Jing, Qichao Wang, Harry H. Cheng, Weiming Peng, and Wenqiang Xu. "A review of vision-based traffic semantic understanding in ITSs." *IEEE Transactions on Intelligent Transportation Systems* (2022).
- [17] Zhang, Xingchen, Yuxiang Feng, Panagiotis Angeloudis, and Yiannis Demiris. "Monocular visual traffic surveillance: A review." *IEEE Transactions on Intelligent Transportation Systems* 23, no. 9 (2022): 14148-14165.
- [18] Wang, Wenguan, Tianfei Zhou, Fatih Porikli, David Crandall, and Luc Van Gool. "A survey on deep learning technique for video segmentation." *arXiv e-prints* (2021): arXiv:2107.
- [19] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III* 18, pp. 234-241. Springer International Publishing, 2015.
- [20] Badrinarayanan, Vijay, Alex Kendall, and Roberto Cipolla. "Segnet: A deep convolutional encoder-decoder architecture for image segmentation." *IEEE transactions on pattern analysis and machine intelligence* 39, no. 12 (2017): 2481-2495.
- [21] Chen, Liang-Chieh, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. "Encoder-decoder with atrous separable convolution for semantic image segmentation." In *Proceedings of the European conference on computer vision (ECCV)*, pp. 801-818. 2018.
- [22] Jha, Debesh, Michael A. Riegler, Dag Johansen, Pål Halvorsen, and Håvard D. Johansen. "Doubleu-net: A deep convolutional neural network for medical image segmentation." In *2020 IEEE 33rd International symposium on computer-based medical systems (CBMS)*, pp. 558-564. IEEE, 2020.
- [23] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778. 2016.
- [24] Zhang, Zhengxin, Qingjie Liu, and Yunhong Wang. "Road extraction by deep residual u-net." *IEEE Geoscience and Remote Sensing Letters* 15, no. 5 (2018): 749-753.
- [25] Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. "Imagenet: A large-scale hierarchical image database." In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248-255. Ieee, 2009.
- [26] Sudre, Carole H., Wenqi Li, Tom Vercauteren, Sebastien Ourselin, and M. Jorge 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: Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings* 3, pp. 240-248. Springer International Publishing, 2017.
- [27] Redmon, Joseph, and Ali Farhadi. "Yolov3: an incremental improvement. 2018." *arXiv preprint arXiv:1804.02767* 20 (1804).
- [28] Zhong, Zhun, Liang Zheng, Guoliang Kang, Shaozi Li, and Yi Yang. "Random erasing data augmentation." In *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 07, pp. 13001-13008. 2020.
- [29] Du, Dawei, Yuankai Qi, Hongyang Yu, Yifan Yang, Kaiwen Duan, Guorong Li, Weigang Zhang, Qingming Huang, and Qi Tian. "The unmanned aerial vehicle benchmark: Object detection and tracking." In *Proceedings of the European conference on computer vision (ECCV)*, pp. 370-386. 2018.
- [30] Lyu, Ye, George Vosselman, Gui-Song Xia, Alper Yilmaz, and Michael Ying Yang. "UAVid: A semantic segmentation dataset for UAV imagery." *ISPRS journal of photogrammetry and remote sensing* 165 (2020): 108-119.
- [31] Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." *arXiv preprint arXiv:1412.6980* (2014).

- [32] Tan, Mingxing, and Quoc Le. "Efficientnet: Rethinking model scaling for convolutional neural networks." In *International conference on machine learning*, pp. 6105-6114. PMLR, 2019.
- [33] Sandler, Mark, Andrew Howard, Menglong Zhu, Andrey Zhmoginov, and Liang-Chieh Chen. "Mobilenetv2: Inverted residuals and linear bottlenecks." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4510-4520. 2018.
- [34] Lin, Tsung-Yi, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. "Feature pyramid networks for object detection." In *Proceedings*.
- [35] Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao. "Yolov4: Optimal speed and accuracy of object detection." *arXiv preprint arXiv:2004.10934* (2020).
- [36] Liu, Wei, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander 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*, pp. 21-37. Springer International Publishing, 2016.



# Strengthening Network Security: Evaluation of Intrusion Detection and Prevention Systems Tools in Networking Systems

Wahyu Adi Prabowo<sup>1</sup>, Khusnul Fauziah<sup>2</sup>, Aufa Salsabila Nahrowi<sup>3</sup>, Muhammad Nur Faiz<sup>4</sup>,  
Arif Wirawan Muhammad<sup>5</sup>

Department of Informatics Engineering-Faculty of Informatics, Institut Teknologi Telkom Purwokerto, Indonesia<sup>1, 2, 3</sup>

Department of Computer and Business, Cybersecurity, Cilacap State Polytechnic, Cilacap, Indonesia<sup>4</sup>

Department of Information Security and Web Technology, Universiti Tun Hussein Onn Malaysia, Johor, Malaysia<sup>5</sup>

**Abstract**—This study aims to enhance network security by comprehensively evaluating various Intrusion Detection and Prevention Systems tools in networking systems. The objectives of this research were to assess the performance of different IDPS tools in terms of computer resources utilization, Quality of Service metrics namely delay, jitter, throughput, and packet loss, and their effectiveness in countering Distributed Denial of Service attacks, specifically ICMP Flood and SYN Flood. The evaluation used popular IDPS tools, including Snort, Suricata, Zeek, OSSEC, and Honeypot Cowrie. Real attack scenarios were simulated to measure the tools performance. The results indicated CPU and RAM usage variations among the tools, with Snort and Suricata showing efficient resource utilization. Regarding QoS metrics, Snort demonstrated superior performance in delay, jitter, throughput, and packet loss mitigation for both attack types. The implication for further research lies in exploring the optimal configurations and fine-tuning of IDPS tools to achieve the best possible network security against DDoS attacks. This research provides valuable insights into selecting appropriate IDPS tools for network administrators, cybersecurity professionals, and organizations to fortify their infrastructure against evolving cyber threats.

**Keywords**—IDPS; network security; computer performance; Quality of Service; DDoS attacks

## I. INTRODUCTION

In today's digital landscape, cybersecurity measures are paramount to protect and protect networks and sensitive data. Among the various cyber threats organizations and individuals face, Distributed Denial of Service (DDoS) attacks pose significant challenges. These attacks involve overwhelming a target system with excessive traffic, rendering it unavailable to legitimate users [1]. Developing effective defense mechanisms against DDoS flooding attacks requires a comprehensive understanding of the problem and the techniques used to prevent, detect, and respond to such attacks [2]. In a DDoS attack, the attacker orchestrates the assault using a network of remotely controlled and widely dispersed nodes. These nodes work collaboratively to flood the victim's network with overwhelming traffic. The primary objective of this attack is not to directly exploit the victim's data but to disrupt the normal functioning of the victim's resources, making it challenging for legitimate users to access the services.

The agent-handler model is a significant structure utilized in DDoS attacks, involving four key participants: the attacker or botmaster, handlers, agents, and the victim [3]. The attacker, also known as the botmaster, communicates indirectly with the agents through the handlers, which act as intermediaries facilitating coordination and communication [4]. The agents compromised devices or systems attack by flooding the victim's network with massive malicious traffic [5], [6].

The agent-handler model provides several advantages for attackers, enabling them to maintain anonymity and distance themselves from the attack [4]. The owners of compromised agent systems often remain unaware that their devices are being exploited to launch DDoS attacks [5]. Moreover, handlers allow the attacker to control multiple agents simultaneously, significantly amplifying the scale and impact of the attack [7]. This model proves particularly effective when targeting web servers during DDoS attacks [5]. By overwhelming the target server with a flood of HTTP requests, such as in an HTTP flood attack, the attacker can exhaust the server's resources, disrupting its availability [7]. Another technique within this model is the Slowloris attack, where partial HTTP requests are sent to the target server, causing it to open additional connections and eventually leading to resource exhaustion [7].

To counter these attacks, Intrusion Detection and Prevention Systems (IDPS) have emerged as a crucial tool in safeguarding networks [8]–[10]. These systems effectively detect and mitigate DDoS attacks to prevent service disruption and data compromise [11]. However, one of the challenges IDPS faces is the ability to effectively detect and defend against evolving and unprecedented attacks [12], [13]. IDPS have traditionally employed two approaches for attack detection: signature-based and anomaly-based [14].

Signature-based detection relies on predefined attack patterns or signatures to identify threats. Although this method can accurately identify known attacks, it becomes ineffective against new or unprecedented attacks that do not match existing signatures [14]. On the other hand, anomaly-based detection analyzes network traffic and identifies abnormal patterns or behavior that deviates from regular network activity [12], [15]. This approach is more effective in detecting unknown attacks, as it does not rely on specific attack

signatures but instead focuses on identifying anomalous behavior [16]. Intrusion detection and prevention system (IDPS) is vital to cybersecurity measures. It is crucial to safeguard computer networks and systems from unauthorized access and malicious activities. IDPS monitors network traffic and analyzes it for any signs of suspicious or malicious behavior [17]–[19]. IDPS includes detecting and preventing unauthorized access attempts, malware attacks, and other security breaches [20], [21].

An IDPS can be either network-based or host-based [22]. Network-based IDPS monitors network traffic at various points in the network infrastructure, such as routers and switches, to identify any abnormal patterns or activities. Host-based IDPS, on the other hand, focuses on monitoring an individual host or endpoint device to detect signs of intrusion or malicious activity [23]. The primary function of an IDPS is to detect and prevent unauthorized access to a network or system. It achieves this by analyzing network traffic and comparing it against predefined patterns or signatures of known attacks [24]. Suppose an IDPS identifies any suspicious activity or a match with a known attack signature. In that case, it generates an alert or takes immediate action to prevent further damage and secure the system [25]. An IDPS can also detect and prevent anomalous behavior not covered by known attack signatures [26].

In the comparative analysis of Intrusion Detection and Prevention Systems (IDPS), several popular systems are evaluated, including Snort, Suricata, Zeek, OSSEC, and the honeypot Cowrie. The existing literature on Intrusion Detection and Prevention Systems (IDPS) is extensive and diverse, with each study providing valuable insights. Based on the comprehensive test results in this study [27], the utilization of pfSense and Suricata emerges as the proposed solution to thwart attacks initiated by internal users and curtail assaults stemming from internal networks, as evidenced by the conducted attack test scenarios. With supplementary devices, the next-generation firewall pfSense and Suricata can significantly bolster network security compared to relying solely on traditional firewalls.

Previous studies have examined the use of IDPS in ensuring Quality of Service in various network environments. These studies have highlighted the importance of IDPS in maintaining network performance and protecting against potential cyber threats. One study was conducted by [28]. Focused on using IDPS in cloud environments to achieve desired security in next-generation networks. The study analyzed different intrusions that could affect cloud resources and services' availability, confidentiality, and integrity. Based on their findings, they recommended positioning IDPS in cloud environments as a crucial step towards ensuring network security. Previous studies have also emphasized the need for IDPS to protect against various attacks, such as distributed denial-of-service attacks, malware infections, and unauthorized access attempts. Another QoS study by [29], [30] also emphasized the role of IDPS in maintaining QoS. Specifically, their study focused on using IDPS in wireless sensor networks. By deploying IDPS in wireless sensor networks, they observed improved QoS metrics such as network reliability, latency, and packet delivery ratio.

Another relevant study, conducted by [31], explores the approach of integrating a Network Intrusion Detection System (NIDS) and a Host-based Intrusion Detection System (HIDS), which can yield more optimal results in addressing security threats. In this approach, Snort is employed as NIDS to detect network-based intrusions by implementing rules capable of recognizing attack patterns. On the other hand, OSSEC functions as HIDS and effectively detects threats at the host level through log analysis, integrity monitoring, and rootkit detection. Both systems complement each other, with NIDS focusing more on network traffic analysis while HIDS concentrates on device and system protection at the host level.

The study by [32] proposes an analytical queuing model for assessing the impact of IDPS performance on network QoS. It explores the trade-off between security and QoS, demonstrating how enhancing security can lead to improved performance, albeit with some trade-offs. The study by [33] employs a multi-objective Bat algorithm to optimize security and QoS in a real-time operating system. It efficiently selects optimal security policies, ensuring minimal disruptions to Quality of Service. These studies offer valuable insights into enhancing network security and QoS through innovative IDPS approaches, highlighting the importance of balancing security measures with network performance considerations.

Contributions from other research, as presented in the studies by [33]–[38], also provide valuable insights within the domain of IDPS. Researchers examine diversity analysis for open-source IDS, aiding security architects in optimizing system performance. The study in [34] proposes a comprehensive multi-cloud integration security framework incorporating honeypots, significantly enhancing attack detection accuracy. The research in [35] introduces SYNGuard, a dynamic threshold-based SYN flood attack detection and mitigation system in Software-Defined Networks (SDNs), and compares the performance of Snort and Zeek IDS. Researchers [36] and [37] present policy-based security configuration management for IDPS, demonstrating its effectiveness using real-world intrusion detection datasets. Meanwhile, [38] analyzes password attacks via honeypots using machine learning techniques to unveil valuable password attack patterns.

Despite the significant insights provided by previous studies regarding the effectiveness and performance of IDPS systems, a comprehensive analysis of Distributed Denial of Service (DDoS) attacks, particularly ICMP Flood and SYN Flood attacks, on networking systems still needs to be improved. This research aims to fill this gap by evaluating the capabilities of IDPS systems such as Snort, Suricata, Zeek, OSSEC, and Honeypot Cowrie within network traffic. Through meticulous experiments, including real attack scenarios and calculations of Quality of Service (QoS) parameters such as throughput, jitter, delay, and packet loss during ICMP Flood and SYN Flood attacks, this study aims to provide valuable insights for network administrators, cybersecurity professionals, and organizations. The ultimate goal is to assist decision-makers in selecting and implementing the most suitable IDPS tools to safeguard their infrastructure against DDoS attacks, particularly in the context of ICMP Flood and SYN Flood attacks.

In the subsequent sections of this paper, the comprehensive analysis of Intrusion Detection and Prevention Systems (IDPS) in the context of DDoS attacks is explored. Following this introduction, the research methodology is described in Section II. Section III presents the results and findings of the experiments, including an evaluation of Snort, Suricata, Zeek, OSSEC, and the honeypot Cowrie. In Section IV, conclusions are provided based on the results and discussions on potential future works to enhance network security further.

## II. METHOD

This research employs an experimental methodology to evaluate the performance of various intrusion detection tools (Snort, Suricata, Zeek, Ossec, and Honeypot Cowrie) in handling specific cyber-attacks, including ICMP Flood and SYN Flood. The research objective is to analyze how each tool responds to the attacks regarding key performance metrics. The independent variables consist of the intrusion detection tools, while the dependent variables include Delay, Jitter, Throughput, and Packet Loss measured before and after the attacks. In Fig. 1, the experimental design encompasses controlled experiments, where each tool is subjected to the same attack scenarios under consistent network conditions.

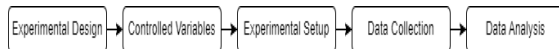


Fig. 1. Research method.

The experimental setup includes deploying the selected tools in a test network environment, and the attacks are initiated to evaluate the detection and response capabilities. The experimental setup involves deploying the selected intrusion detection tools within a controlled test network environment. Subsequently, targeted cyber-attacks are initiated to rigorously evaluate and assess each tool's detection and response capabilities. This evaluation allows for a comprehensive analysis of their performance under realistic attack scenarios, providing valuable insights into their effectiveness in safeguarding computing systems against potential threats.

Data collection involves meticulously recording each tool's performance metrics during the attack simulations. Throughout the simulations, relevant performance data, including Delay, Jitter, Throughput, and Packet Loss, is carefully documented for each detection tool. Quality of Service (QoS) is a method used to measure the quality of a network and determine the level of service it provides. QoS measures specific performance characteristics such as Delay, Jitter, Throughput, and Packet Loss, which are associated with a service [39], [40].

1) *Throughput*: Throughput refers to the actual bandwidth measured at a specific time when sending a file. Unlike bandwidth, which is measured in bits per second (bps), throughput better represents the actual bandwidth at a specific moment and under certain network conditions, particularly when downloading a particular file. It is calculated as the total number of successfully transmitted data (in bits) divided by the total time taken to transmit that data (in seconds):

$$\text{Throughput} = \frac{\text{Total amount of transmitted data}}{\text{Total time to transmit the data}} \quad (1)$$

2) *Packet loss*: Packet Loss is the percentage of packets lost during data transmission. Various factors, such as weak signals in the network, network hardware errors, or environmental interference, can cause this. Packet Loss is a critical parameter that illustrates the number of lost packets due to collisions and congestion in the network. It is calculated as follows:

$$\text{Packet Loss} = \frac{\text{Number of lost packets}}{\text{Total Number of packets sent}} \times 100\% \quad (2)$$

3) *Jitter*: Jitter is the variation in delay (time difference) between packets in the network, which is influenced by the queue length when processing data. It is affected by the traffic load and the number of packets (congestion) in the network, particularly during periods of high traffic. Jitter is calculated using the following equation:

$$\text{Jitter} = \frac{\text{Total delay variation}}{\text{Total amount of transmitted data}} \quad (3)$$

4) *Delay*: Delay or Latency is the time it takes for data to travel from the source to the destination. The delay is influenced by distance, physical media, congestion, and processing times. It is calculated as follow:

$$\text{Delay} = \frac{\text{Total delay}}{\text{Total amount of transmitted data}} \quad (4)$$

In cybersecurity, particularly in defending against Distributed Denial of Service (DDoS) attacks, IDPS plays a pivotal role. To bolster the effectiveness of IDPS in countering the ever-evolving DDoS threats, it becomes imperative to incorporate more analytical metrics. One such metric that merits heightened attention is the Detection Rate (DR) [41], calculated as follows:

$$\text{Detection Rate} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (5)$$

The Detection Rate (DR) is a critical metric that gauges the system's ability to identify genuine DDoS attacks accurately among all positive instances. True Positives (TP) represent instances where the IDPS correctly identifies and labels a legitimate DDoS attack. At the same time, False Negatives (FN) indicates instances where the system fails to detect a real DDoS threat, potentially leading to a security breach. In the DDoS mitigation landscape, the significance of DR cannot be overstated. It is a cornerstone for evaluating the IDPS aptitude to identify and thwart DDoS attacks precisely. Achieving a high DR is paramount as it minimizes the risk of false negatives, ensuring that legitimate DDoS threats do not go undetected.

By adopting this approach, this research acquires comprehensive and detailed data on the performance of each detection tool under various attack scenarios. Subsequently, data analysis entails statistical comparisons to determine significant differences in performance metrics between the tools. The data analysis process encompasses conducting thorough statistical comparisons to discern notable variations in performance metrics among the different detection tools. Through the application of advanced statistical techniques, the aim is to identify any statistically significant differences in the

performance of each tool. This rigorous analysis enables us to gain valuable insights into the relative strengths and weaknesses of the detection tools, facilitating a comprehensive assessment of their capabilities in handling diverse cyber-attacks.

The rigorous experimental methodology aims to provide reliable insights into the efficiency and effectiveness of intrusion detection tools in diverse computing environments, particularly under varying attack conditions. Through comprehensive evaluations and controlled experiments, valuable data is sought to assess the capabilities and performance of these tools in safeguarding computing systems against a wide range of potential cyber threats. Doing so aims to establish a robust understanding of IDPS Tools, enhance cybersecurity practices, and ensures a more secure computing landscape.

### III. RESULT AND FINDING

#### A. Experimental Design

The experimental design employed in this study involved conducting controlled experiments to evaluate the performance of each intrusion detection tool under consistent network conditions. All selected tools were subjected to the same attack scenarios in a controlled test network environment to ensure a fair and unbiased assessment. For the attack scenario in Fig. 2, the researchers utilized a computer laboratory comprising ten computers infiltrated with DDoS bots controlled by an attacker operating the handler. This simulation was used to launch attacks on a server, from which the necessary data was obtained during the testing of IDPS (Intrusion Detection and Prevention System) tools, including Snort, Suricata, Zeek, Ossec, and Honeypot Cowrie. This simulation aimed to assess the IDPS tools' performance in detecting and responding to the DDoS attacks orchestrated by the attacker through the compromised bots. The data collected from these simulated attacks served as crucial input for evaluating and analyzing the effectiveness of each IDPS tool in defending against such cyber threats.

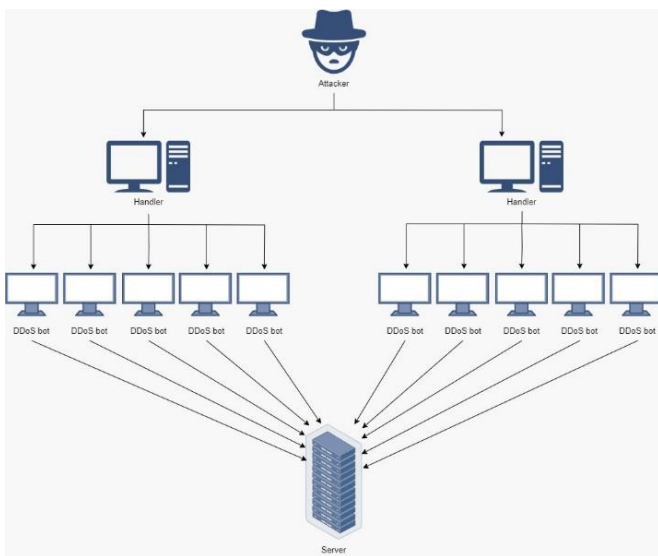


Fig. 2. Attack scenario.

Through controlled experiments, the aim was to eliminate any potential confounding variables and ensure that the observed differences in performance metrics were solely attributed to the capabilities of the intrusion detection tools. Each tool underwent testing under identical conditions, including network traffic, attack intensity, and duration. This standardized approach allowed for objectively comparing the tools' performance and drawing meaningful conclusions about their efficacy in detecting and mitigating various cyber-attacks. The performance metrics, such as RAM usage, CPU utilization, network throughput, delay, jitter, and packet loss, were carefully monitored and recorded during the attack simulations for each tool. Furthermore, to enhance the reliability of the findings, the experiments were repeated multiple times to account for any random variations and ensure the consistency of the results. The aggregated data from the repeated experiments provided a more robust basis for analysis and interpretation.

#### B. Controlled Variables

A carefully selected set of hardware specifications was strategically employed to ensure the successful acquisition of pertinent data for the research. These specifications were pivotal in establishing a robust experimental environment, enabling controlled experiments and the meticulous recording of performance metrics for the intrusion detection tools under investigation. With utmost attention to detail, specific hardware components were carefully chosen and implemented, tailored precisely to align with the research objectives. The following hardware specifications in Table I were utilized to facilitate data collection.

TABLE I. EXPERIMENTAL HARDWARE TOOLS

No	Hardware	Version	Number of Tools	Ip Number
1	switch	cisco sf95d-16 16-port 10/100	2 unit	192.168.100.150 & 192.168.100.151
2	computer server	server dell t150 xeon e-2324g	1 unit	192.168.100.154
3	computer server idps	server dell t40 xeon e-2224g	1 unit	192.168.100.153
4	computer idps console	all in one (aio) pc dell optiplex 7440	1 unit	192.168.100.152
5	computer agent	asus pc all in one v222gak wa141t - dualcore	10 unit	192.168.100.1-10
6	computer handler	asus pc all in one v222gak wa141t - dualcore	2 unit	192.168.100.11 & 192.168.100.12
7	computer attacker	hp pavilion aero 13 be2001au ryzen 5 7535u	1 unit	192.168.100.13

Meticulously designed and implemented a network topology for this research, as illustrated in Fig. 3, which comprised a carefully selected set of computers, each assigned specific roles. At the heart of the topology, the computer server served as a centralized repository for data. At the same time, the deployment of IDPS tools spanned across multiple computers, including servers, effectively safeguarding the network traffic from potential DDoS attacks. The assignment

of IP addresses was skillfully managed through the switch, distributing the network across 13 computers. Among these designated systems, ten were dedicated to functioning as agent botnets for DDoS, two served as handlers with control over the agents, and one acted as the attacker. This research opted for the Kali Linux 2023.1 operating system, facilitating the smooth integration of essential intrusion detection tools, namely Snort, Suricata, Zeek, Ossec, and Honeypot Cowrie. This research employed Wireshark 4.0 as the chosen monitoring tool to ensure efficient network traffic monitoring.

The hardware setup and carefully crafted network topology laid the foundation for the controlled experiments, enabling us to systematically assess the performance of each intrusion detection tool under varying attack scenarios. By employing a standardized approach, reliability and accuracy in research results were ensured, providing the means to make informed evaluations regarding the capabilities and effectiveness of these tools in countering diverse cyber threats.

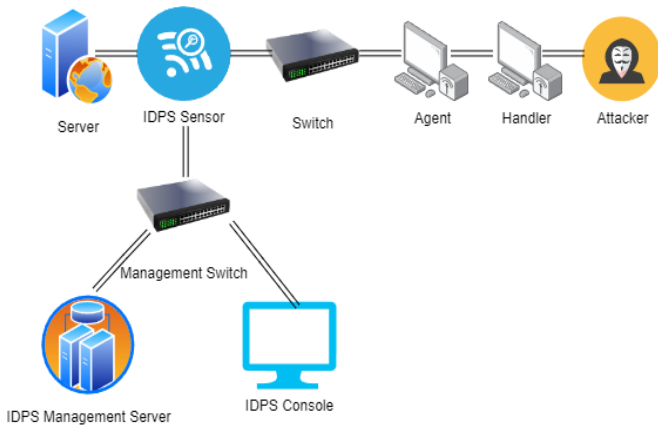


Fig. 3. Network topology.

### C. Experimental Setup

The experimental Setup section of this research focuses on the systematic deployment and evaluation of several intrusion detection tools, namely Snort, Suricata, Zeek, Ossec, and Honeypot Cowrie. Each tool is selected individually and installed with its respective configurations. Subsequently, comprehensive testing assesses their performance in handling DDoS attacks, specifically through ICMP Flood and TCP SYN Flood.

Initiate the evaluation process, the server is configured with rules specific to each IDPS tool, and simulated attacks are launched from an attacker's PC to the server using the DDoS tool Hping3. The commands for the SYN Ddos attack and Icmp Dodos attack simulations are provided as follows in Fig. 4 and Fig. 5:

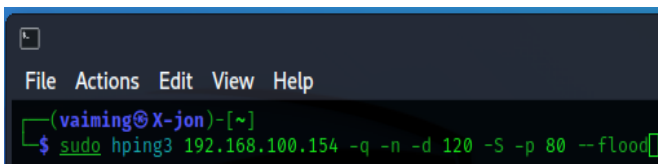


Fig. 4. SYN DDoS attack.

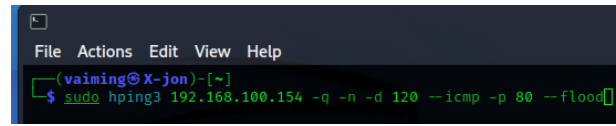


Fig. 5. ICMP DDoS attack.

The server's response to the attack is initially observed when protected by the Snort tool with the IPS command. The objective is to determine whether Snort can successfully detect and generate warnings for the simulated attacks. Once the results are obtained from the Snort testing, the same evaluation process is repeated using the other selected tools, Suricata, Zeek, Ossec, and Honeypot Cowrie, on the server.

### D. Data Collection

IDPS like Suricata, Zeek, Ossec, and Honeypot Cowrie play pivotal roles in safeguarding digital environments from malicious activities. They operate as the first line of defense, tirelessly monitoring network traffic and system logs. To assess the efficacy of these systems, metrics like the Detection Rate (DR) are paramount. In the context of this research, the Detection Rate (DR) emerges as a pivotal metric in assessing the performance of the IDPS. With 128,027 True Positives, signifying accurate identifications of actual intrusion attempts, and a relatively low 4,241 False Negatives (FN), which represent instances where genuine threats were not detected, the IDPS demonstrates a robust capability in effectively distinguishing malicious activities from benign network traffic. The DR, calculated as the ratio of True Positives to the sum of True Positives and False Negatives ( $TP / (TP + FN)$ ), reflects the system's ability to capture a high proportion of genuine intrusions. This value is a crucial requirement in this research, where achieving a DR of 128,027, or 97% of all intrusion attempts, is integral to minimizing the risk of false negatives and ensuring the thorough protection of digital assets.

In Table II and Table III, the performance of each IDPS tool was carefully observed during the ICMP flood attack. Snort exhibited a slight increase in RAM usage by 0.07%, followed by a slightly larger increase in CPU usage by 5.00%. Conversely, Suricata experienced a more pronounced rise in RAM usage by 0.19% and a substantial increase in CPU usage by 16.67%. Zeek demonstrated minimal fluctuations in RAM and CPU usage, with only 0.09% and 0.00% changes, respectively. OSsec recorded a moderate uptick in RAM and CPU usage, showing increases of 0.08% and 2.70%, respectively, highlighting its ability to manage ICMP flood attacks without significant overhead.

In contrast, Honeypot Cowrie displayed a noticeable increase in RAM usage by 0.13%, followed by a slightly more substantial rise in CPU usage of 3.61%. Network performance during the ICMP flood attack revealed diverse trends. Snort indicated a moderate upswing in network throughput, measuring 578.79 kb/s. Conversely, Suricata, Zeek, and Ossec experienced slight decreases in network throughput by 1.13 kb/s, 0.66 kb/s, and 0.05 kb/s, respectively. Remarkably, honeypot cowrie showcased a significant spike in network throughput, reaching 701.07 kb/s, underscoring its efficiency in addressing ICMP flood attacks.

TABLE II. COMPUTER PERFORMANCE BEFORE ATTACK

IDPS Tools	DDoS Attack	Before Attack		
		Ram (%)	Cpu (%)	Network (kb/s)
snort	icmp flood	27.30	0.17	32.38
	syn flood	27.56	0.22	36.4
suricata	icmp flood	32.19	0.66	34.39
	syn flood	36.26	0.82	35.40
zeek	icmp flood	32.69	0.22	32.78
	syn flood	32.57	0.52	32.61
ossec	icmp flood	32.57	0.37	32.70
	syn flood	33	0.22	32.65
honeypot cowrie	icmp flood	29.75	0.415	32.65
	syn flood	31.91	0.52	35.90

TABLE III. COMPUTER PERFORMANCE AFTER ATTACK

IDPS Tools	DDoS Attack	After Attack		
		Ram (%)	Cpu (%)	Network (kb/s)
snort	icmp flood	27.37	5.17	611.17
	syn flood	27.74	8.11	3238.76
suricata	icmp flood	32.38	7.67	856.26
	syn flood	36.4	9.02	3666.62
zeek	icmp flood	32.78	3.12	983.34
	syn flood	32.61	6.07	3569.07
ossec	icmp flood	32.65	8.08	566.89
	syn flood	32.39	7.73	3,596.54
honeypot cowrie	icmp flood	29.88	6.42	733.72
	syn flood	32.07	8.57	3452.69

Turning to the SYN Flood attack, the IDPS tools once again exhibited diverse patterns of performance adjustment. Snort displayed a slight increase in RAM usage by 0.18%, followed by a more substantial rise in CPU usage by 27.27%. Suricata showcased a more significant uptick in RAM usage by 0.74% and a noteworthy increase in CPU usage of 17.07%. Zeek demonstrated minimal RAM and CPU usage variations, with only 0.09% and 0.00% changes, respectively. Conversely, OSSEC recorded slightly decreased RAM usage by 0.22%, while CPU usage increased by 0.00%. HoneyPot Cowrie experienced a noticeable increase in RAM usage by 0.13% and a relatively significant rise in CPU usage of 6.59%.

Network performance during the SYN Flood attack also revealed distinct behavior. Snort and Suricata exhibited moderate increases in network throughput, measuring 2205.59 KB/s and 1587.86 KB/s, respectively, demonstrating their efficient responses to SYN Flood attacks. Zeek demonstrated a slight decrease in network throughput by 1.34 KB/s, while OSSEC experienced a significant surge in network throughput, reaching 1929.89 KB/s. Notably, honeyPot cowrie significantly increased network throughput, measuring 2413.52 KB/s, further highlighting its robustness in handling SYN Flood attacks.

These observations suggest differences in the tools' ability to detect and counter such attacks. In the case of ICMP Flood attacks, it was observed that certain IDPS tools experienced

notable increases in resource utilization, such as RAM, CPU, and network throughput, after the attacks. These observations imply varying sensitivity and adaptability of these tools to the attack type. Similarly, during SYN Flood attacks, the IDPS tools exhibited diverse patterns of resource usage alterations, suggesting differences in their ability to detect and counter such attacks. The observed changes in performance metrics underscore the need for a nuanced evaluation of IDPS tools under distinct attack scenarios.

This study in Fig. 6 conducted QoS measurements for throughput during ICMP Flood and SYN Flood DDoS attacks using different Intrusion Detection and Prevention Systems (IDPS) tools, namely Snort, Suricata, Zeek, Ossec, and HoneyPot. The results indicated variations in throughput values across these tools for both attack types. Among the tested tools, Snort demonstrated the highest throughput during ICMP Flood attacks, reaching 26,485 bits per second. At the same time, Suricata and Zeek showed similar throughput values at 32,400 and 32,438 bits per second, respectively. Ossec and HoneyPot yielded slightly lower throughputs at 26,052 and 39,897 bits per second, respectively.

For SYN Flood attacks, Snort exhibited the highest throughput of 29,701 bits per second, followed closely by HoneyPot at 34,701 bits per second. Suricata and Zeek yielded lower throughput values at 25,029 and 21,970 bits per second, respectively. Ossec demonstrated the lowest throughput among the tested tools for SYN Flood attacks, registering 21,970 bits per second.

Snort exhibited strong throughput values for both ICMP Flood and SYN Flood attacks, making it a viable choice for mitigating these attack types. Suricata and Zeek also demonstrated competitive throughput, indicating their potential effectiveness in handling DDoS attacks.

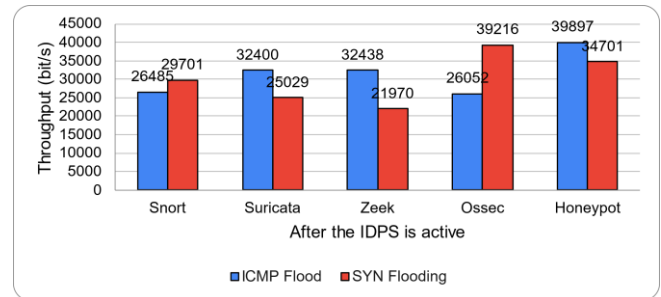


Fig. 6. Throughput DDoS attack.

The Delay values for different types of DDoS attacks were evaluated using various Intrusion Detection and Prevention Systems (IDPS), including Snort, Suricata, Zeek, Ossec, and HoneyPot Cowrie in Fig. 7. For ICMP Flood attacks, Snort exhibited a delay of 223.53 ms, Suricata had a delay of 183.85 ms, Zeek showed a delay of 45.59 ms, Ossec had a delay of 130.9 ms, and HoneyPot Cowrie displayed the lowest delay of 22.88 ms. Similarly, for SYN Flood attacks, Snort demonstrated a delay of 187.17 ms, Suricata had a delay of 104.59 ms, Zeek exhibited a delay of 17.9 ms, Ossec showed a delay of 60.8 ms, and HoneyPot Cowrie had a delay of 187.2 ms. These Delay values provide insights into the responsiveness of each IDPS in detecting and mitigating ICMP

and SYN Flood attacks. It is worth noting that Honeypot Cowrie consistently displayed lower Delay values, indicating its potential effectiveness in handling such attacks with minimal delay.

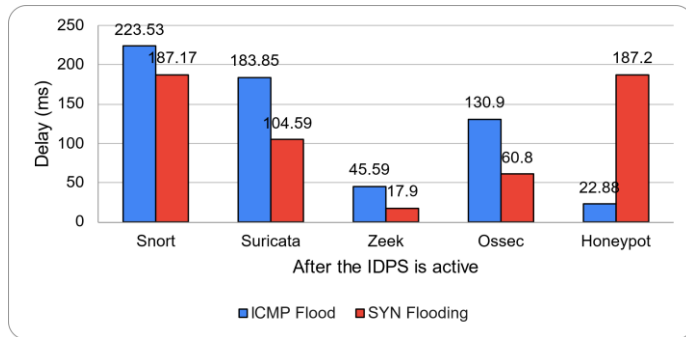


Fig. 7. Delay DDoS attack.

Analyzing jitter values across various DDoS attack scenarios and corresponding Intrusion Detection and Prevention Systems (IDPS) tools reveals distinct patterns in Fig. 8. In the case of ICMP Flood attacks, Zeek stands out with remarkably low jitter (0.88 ms), indicating stable and consistent packet delay. Conversely, Snort (7.37 ms), Suricata (1.8 ms), Ossec (1.01 ms), and Honeypot (11.5 ms) exhibit comparatively higher jitter values, suggesting potential fluctuations in delay times. A similar trend emerges during SYN Flooding attacks, where Zeek maintains its superior performance in jitter control (2.02 ms). Suricata (5.63 ms) and Ossec (6.08 ms) demonstrate increased jitter, while Snort (1.81 ms) and Honeypot (1.82 ms) exhibit relatively better control. These findings underscore Zeek's consistent jitter management capabilities across both attack types.

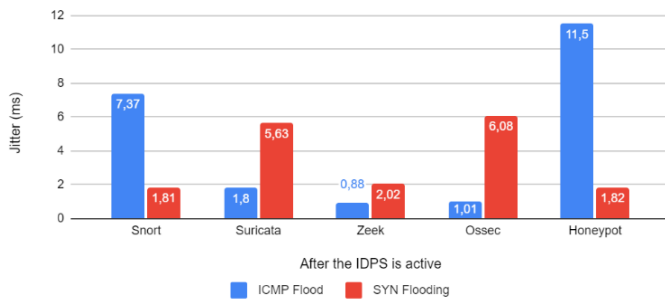


Fig. 8. Jitter DDoS attack.

The investigation into packet loss rates during ICMP Flood and SYN Flooding attacks, evaluated across a range of Intrusion Detection and Prevention Systems (IDPS) tools, yielded distinct outcomes. In Figure 9, Snort and Suricata exhibited minimal packet loss, recording percentages of 0.32% and 0.44% for ICMP Flood and 0.56% and 0.29% for SYN Flooding, respectively. Zeek displayed effective packet loss mitigation, with rates of 0.25% for ICMP Flood and 0.14% for SYN Flooding. Ossec and Honeypot Cowrie demonstrated slightly higher packet loss percentages, at 0.33% and 0.19% for ICMP Flood and 0.56% for SYN Flooding. These findings illuminate the diverse packet loss responses of IDPS tools to specific attack scenarios, empowering network administrators

and cybersecurity practitioners with valuable insights for optimizing DDoS protection strategies.

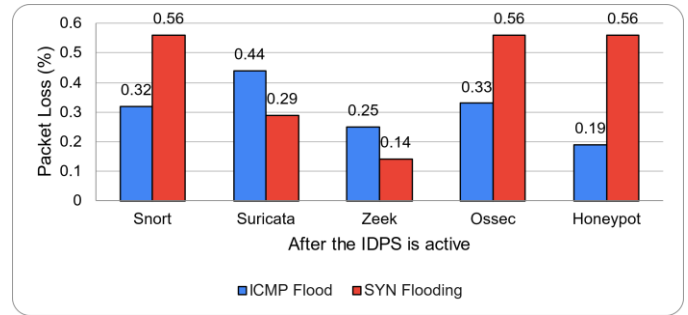


Fig. 9. Packet loss DDoS attack.

The Quality of Service (QoS) analysis of the network performance before and after different attack scenarios, as measured by various metrics, offers valuable insights into the effectiveness of the Intrusion Detection and Prevention Systems (IDPS) tools. Among the tested tools, Snort and Zeek consistently demonstrate a relatively robust ability to mitigate the impact of attacks on network Delay and Packet Loss. Suricata and OSSEC, on the other hand, exhibit more susceptibility to disruptions caused by the attacks, with increased Delay, Jitter, and Packet Loss, especially evident in SYN Flood attacks. Notably, Honeypot Cowrie proves adept at maintaining network stability during ICMP Flood attacks, showcasing lower Jitter and relatively stable Throughput. These observations underline the varying QoS responses of different IDPS tools to distinct attack types, providing crucial insights for making informed decisions regarding network defense strategies. The ICMP Flood attacks generally result in increased Delay and Packet Loss, while the SYN Flood attacks tend to affect Delay, Jitter, and sometimes throughput. Among the IDPS tools, Snort and Zeek exhibit relatively better network performance maintenance, while Suricata and OSSEC show more impact from the attacks. Honeypot Cowrie maintains network performance relatively well, particularly for ICMP Flood attacks. It is important to note that these observations provide insights into how each IDPS tool responds to specific attack types regarding QoS metrics.

### E. Data Analysis

The data analysis phase serves as the foundation of the investigation, shedding light on the performance dynamics of distinct intrusion detection tools when confronted with diverse cyber threats. This research systematically compared key performance metrics through meticulous experimentation and keen observation before and after simulated attacks. The analytical focus encompassed critical parameters, including delay, jitter, throughput, and packet loss, offering a comprehensive view of each tool's response. Notably, Snort exhibited commendable efficiency in managing ICMP Flood attacks, showcasing minimal network latency and jitter disruption. Suricata demonstrated adeptness in mitigating SYN Flood attacks with modest fluctuations. Zeek's proficiency shone through its stable network throughput during ICMP Flood scenarios.

Meanwhile, OSSEC displayed a robust defense mechanism against ICMP Flood attacks, containing packet loss within

acceptable bounds. HoneyPot Cowrie effectively mitigated packet loss while experiencing elevated jitter during ICMP Flood incidents. The analysis, bolstered by robust statistical techniques, revealed nuanced performance differences allowing us to conclude each tool's strengths and limitations. These insights offer essential guidance for practitioners configuring intrusion detection tools advancing cybersecurity defense strategies with precision.

This research analysis of countering ICMP DDoS and SYN Flood attacks highlighted varying degrees of efficacy among the IDPS tools. Snort stood out in addressing ICMP Flood attacks, effectively minimizing network disruption, latency, and jitter. Similarly, Suricata exhibited proficiency in mitigating SYN Flood attacks, maintaining stable network throughput, and responding to anomalous traffic patterns. On the other hand, Zeek displayed commendable network throughput during ICMP Flood scenarios but showed moderate fluctuation against SYN Flood attacks. While OSSEC contained packet loss during ICMP Flood incidents, it faced challenges maintaining network stability under SYN Flood onslaughts. HoneyPot Cowrie, resilient against packet loss, experienced elevated jitter during ICMP Flood attacks. These findings collectively suggest that Snort and Suricata are potent contenders for countering ICMP DDoS and SYN Flood attacks, offering consistent and robust responses. The nuanced strengths and limitations underscore the importance of tailored tool selection based on the specific threat landscape and operational requirements.

A comprehensive analysis of QoS data and computer/networking performance metrics reveals that Snort is a standout performer in countering ICMP Flood attacks. This conclusion is drawn from consistent and commendable results across various parameters. Snort effectively mitigated delays and jitter, ensuring optimal network responsiveness and maintaining impressive throughput levels, all while demonstrating minimal packet loss. Moreover, Snort efficiently utilized CPU and RAM resources, indicating its ability to handle ICMP Flood attacks without overstraining the system. These findings position Snort as the most robust IDPS tool for effectively countering ICMP Flood attacks, making it a compelling choice for defending against such threats and ensuring network stability and performance.

Similarly, the analysis indicates that Zeek is the most effective IDPS tool for countering SYN Flood attacks. Zeek consistently demonstrated remarkable performance in minimizing delays and jitter during SYN Flood attacks, maintaining stable network responsiveness. Additionally, Zeek maintained competitive throughput levels and remarkably low packet loss, showcasing its proficiency in managing SYN requests. From a computer and networking performance standpoint, Zeek efficiently allocated CPU and RAM resources, indicating its capability to handle SYN Flood attacks without burdening the system. Overall, Zeek's strong performance across QoS metrics and resource management makes it the optimal choice for countering SYN Flood attacks and safeguarding network stability.

In this comparative analysis of the results, we have examined this research alongside relevant previous studies.

The study by [28], [35] introduces an analytical model for assessing IDPS configurations, emphasizing theoretical modeling. In contrast, the results of this research delve into practical IDPS implementation within a networking system environment to defend against specific threats, emphasizing real-world application. The studies cited as [11], [12], [29], [30], [33], [36]–[38], on the other hand, differ significantly from this research outcome. Given these variations in goals and approaches, direct result comparisons can be challenging. This study's results highlight practical implementation and threat defense, distinguishing it from theoretical modeling and the differing contexts in previous studies.

#### IV. CONCLUSION AND FUTURE WORKS

This study undertook a comprehensive analysis of diverse Intrusion Detection and Prevention Systems (IDPS) tools, namely Snort, Suricata, Zeek, OSSEC, and HoneyPot Cowrie, with a primary focus on their effectiveness in countering Distributed Denial of Service (DDoS) attacks. Through a meticulous evaluation encompassing aspects of network traffic analysis, Quality of Service (QoS) metrics, computer performance, and attack mitigation, this research gained insights into the capabilities of these tools. In this assessment, research revealed distinct performance characteristics for each IDPS tool. Snort excelled in network-based intrusion detection, efficiently identifying and countering threats at the network level. Suricata demonstrated prowess in packet processing and rule matching, making it a strong contender for network security. With its emphasis on comprehensive traffic analysis, Zeek offered valuable insights into network activity. OSSEC showcased robust host-based intrusion detection capabilities, providing effective log analysis and threat identification. HoneyPot Cowrie displayed potential while highlighting areas for improvement in QoS metrics and computer performance. Regarding Quality of Service (QoS), the analysis unveiled Snort as the most effective IDPS tool in countering ICMP Flood and SYN Flood attacks, consistently exhibiting superior throughput, lower delay, minimal jitter, and commendable packet loss rates. These QoS metrics reflect Snort's adeptness in preserving network integrity and minimizing disruption during DDoS incidents.

Future research avenues include integrating advanced machine learning techniques into IDPS tools to optimize detection accuracy while minimizing false positives. Additionally, exploring the deployment of IDPS in dynamic cloud and hybrid environments, understanding their scalability, and adapting them to varying network conditions would provide valuable insights. In conclusion, this study provides valuable insights into the performance of diverse IDPS tools against DDoS attacks. By addressing identified limitations and pursuing avenues for future research, this research can advance the field of network security and contribute to developing resilient defense mechanisms against evolving cyber threats.

#### REFERENCES

- [1] W. A. Al-Khater, S. Al-Maadeed, A. A. Ahmed, A. S. Sadiq, and M. K. Khan, "Comprehensive review of cybercrime detection techniques," *IEEE Access*, vol. 8, 2020, doi: 10.1109/ACCESS.2020.3011259.
- [2] S. T. Zargar, J. Joshi, and D. Tipper, "A survey of defense mechanisms against distributed denial of service (DDoS) flooding attacks," *IEEE*



- Communications Surveys and Tutorials, vol. 15, no. 4, 2013, doi: 10.1109/SURV.2013.031413.00127.
- [3] N. Hoque, D. K. Bhattacharyya, and J. K. Kalita, "Botnet in DDoS Attacks: Trends and Challenges," *IEEE Communications Surveys and Tutorials*, vol. 17, no. 4, 2015, doi: 10.1109/COMST.2015.2457491.
- [4] M. Masdari and M. Jalali, "A survey and taxonomy of DoS attacks in cloud computing," *Security and Communication Networks*, vol. 9, no. 16, 2016. doi: 10.1002/sec.1539.
- [5] E. Alomari, S. Manickam, B. B. Gupta, S. Karuppayah, and R. Alfaris, "Botnet-based Distributed Denial of Service (DDoS) Attacks on Web Servers: Classification and Art," *Int J Comput Appl*, vol. 49, no. 7, 2012, doi: 10.5120/7640-0724.
- [6] S. Specht and R. Lee, "Taxonomies of Distributed Denial of Service Networks, Attacks, Tools, and Countermeasures," 2003. [Online]. Available: [www.princeton.edu](http://www.princeton.edu)
- [7] M. K. Kareem, O. D. Aborisade, S. A. Onashoga, T. Sutikno, and O. M. Olayiwola, "Efficient model for detecting application layer distributed denial of service attacks," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 1, 2023, doi: 10.11591/eei.v12i1.3871.
- [8] M. Shaohui, G. Tuerhong, M. Wushouer, and T. Yibulayin, "PCA mix-based Hotelling's T2 multivariate control charts for intrusion detection system," *IET Inf Secur*, vol. 16, no. 3, 2022, doi: 10.1049/ise2.12051.
- [9] P. Sai Chowdary and D. Vinod, "Host Intrusion Detection System Using Novel Predefined Signature Patterns by Comparing Random Forest over Decision Tree Algorithm," in *Advances in Parallel Computing*, 2022. doi: 10.3233/APC220092.
- [10] J. Gabirondo-Lopez, J. Egana, J. Miguel-Alonso, and R. Orduna Urrutia, "Towards Autonomous Defense of SDN Networks Using MuZero Based Intelligent Agents," *IEEE Access*, vol. 9, 2021, doi: 10.1109/ACCESS.2021.3100706.
- [11] K. Alsubhi and H. M. AlJadhali, "Intrusion detection and prevention systems as a service in cloud-based environment," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 7, 2018, doi: 10.14569/IJACSA.2018.090738.
- [12] A. H. B. Aighuraibawi, R. Abdullah, S. Manickam, and Z. A. A. Alyasseri, "Detection of ICMPv6-based DDoS attacks using anomaly based intrusion detection system: A comprehensive review," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 6, 2021. doi: 10.11591/ijece.v11i6.pp5216-5228.
- [13] S. Laqtib, K. El Yassini, and M. L. Hasnaoui, "A technical review and comparative analysis of machine learning techniques for intrusion detection systems in MANET," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 3, 2020. doi: 10.11591/ijece.v10i3.pp2701-2709.
- [14] P. Araujo et al., "Impact of Feature Selection Methods on the Classification of DDoS Attacks using XGBoost," *Journal of Communication and Information Systems*, vol. 36, no. 1, 2021, doi: 10.14209/jcis.2021.22.
- [15] N. Z. M. Safar, N. Abdullah, H. Kamaludin, S. A. Ishak, and M. R. M. Isa, "Characterising and detection of botnet in P2P network for UDP protocol," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 18, no. 3, 2020, doi: 10.11591/ijeecs.v18.i3.pp1584-1595.
- [16] C. Oh, J. Ha, and H. Roh, "A survey on tls-encrypted malware network traffic analysis applicable to security operations centers," *Applied Sciences (Switzerland)*, vol. 12, no. 1, 2022, doi: 10.3390/app12010155.
- [17] I. Mukhopadhyay, M. Chakraborty, and S. Chakrabarti, "A Comparative Study of Related Technologies of Intrusion Detection & Prevention Systems," *Journal of Information Security*, vol. 02, no. 01, 2011, doi: 10.4236/jis.2011.21003.
- [18] M. Ozkan-Okay, R. Samet, O. Aslan, and D. Gupta, "A Comprehensive Systematic Literature Review on Intrusion Detection Systems," *IEEE Access*, vol. 9, 2021. doi: 10.1109/ACCESS.2021.3129336.
- [19] A. S. Putra and N. Surantha, "Internal threat defense using network access control and Intrusion Prevention System," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 9, 2019, doi: 10.14569/ijacsa.2019.0100948.
- [20] M. H. Kamarudin, C. Maple, T. Watson, and N. S. Safa, "A New Unified Intrusion Anomaly Detection in Identifying Unseen Web Attacks," *Security and Communication Networks*, vol. 2017, 2017, doi: 10.1155/2017/2539034.
- [21] M. Ahsan, K. E. Nygard, R. Gomes, M. M. Chowdhury, N. Rifat, and J. F. Connolly, "Cybersecurity Threats and Their Mitigation Approaches Using Machine Learning—A Review," *Journal of Cybersecurity and Privacy*, vol. 2, no. 3, 2022, doi: 10.3390/jcp2030027.
- [22] V. Vasilyev and R. Shamsutdinov, "Distributed Intelligent System of Network Traffic Anomaly Detection Based on Artificial Immune System," 2019. doi: 10.2991/itids-19.2019.7.
- [23] B. Hameed, A. AlHabsby, and K. Eldahshan, "Distributed Intrusion Detection Systems in Big Data: A Survey," *Al-Azhar Bulletin of Science*, vol. 32, no. 1, 2021, doi: 10.21608/absb.2021.63810.1100.
- [24] O. Alkadi, N. Moustafa, and B. Turnbull, "A Review of Intrusion Detection and Blockchain Applications in the Cloud: Approaches, Challenges and Solutions," *IEEE Access*, vol. 8, 2020. doi: 10.1109/ACCESS.2020.2999715.
- [25] T. Andrysiak, Ł. Saganowski, and W. Mazurczyk, "Network anomaly detection for railway critical infrastructure based on autoregressive fractional integrated moving average," *EURASIP J Wirel Commun Netw*, vol. 2016, no. 1, 2016, doi: 10.1186/s13638-016-0744-8.
- [26] I. Singh, S. Singhal, and V. Kumar, "Database intrusion detection using role and user level sequential pattern mining and fuzzy clustering," *International Journal of Engineering Research and Technology*, vol. 13, no. 6, 2020, doi: 10.37624/ijert/13.6.2020.1173-1178.
- [27] A. J. Alhasan and N. Surantha, "Evaluation of Data Center Network Security based on Next-Generation Firewall," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 9, 2021, doi: 10.14569/IJACSA.2021.0120958.
- [28] C. Modi, D. Patel, B. Borisaniya, H. Patel, A. Patel, and M. Rajarajan, "A survey of intrusion detection techniques in Cloud," *Journal of Network and Computer Applications*, vol. 36, no. 1, 2013. doi: 10.1016/j.jnca.2012.05.003.
- [29] C. Birkinshaw, E. Rouka, and V. G. Vassilakis, "Implementing an intrusion detection and prevention system using software-defined networking: Defending against port-scanning and denial-of-service attacks," *Journal of Network and Computer Applications*, vol. 136, 2019, doi: 10.1016/j.jnca.2019.03.005.
- [30] H. Hendrawan, P. Sukarno, and M. A. Nugroho, "Quality of service (QoS) comparison analysis of snort IDS and Bro IDS application in software define network (SDN) architecture," in *2019 7th International Conference on Information and Communication Technology, ICoICT 2019*, 2019. doi: 10.1109/ICoICT.2019.8835211.
- [31] D. W. Y. O. Waidyarathna, W. V. A. C. Nayantha, W. M. T. C. Wijesinghe, and K. Y. Abeywardena, "Intrusion Detection System with correlation engine and vulnerability assessment," *International Journal of Advanced Computer Science and Applications*, vol. 9, no. 9, 2018, doi: 10.14569/ijacsa.2018.090947.
- [32] K. Alsubhi, M. F. Zhani, and R. Boutaba, "Embedded Markov process based model for performance analysis of Intrusion Detection and Prevention Systems," in *GLOBECOM - IEEE Global Telecommunications Conference*, 2012. doi: 10.1109/GLOCOM.2012.6503227.
- [33] H. Asad and I. Gashi, "Dynamical analysis of diversity in rule-based open source network intrusion detection systems," *Empir Softw Eng*, vol. 27, no. 1, 2022, doi: 10.1007/s10664-021-10046-w.
- [34] T. Alyas et al., "Multi-Cloud Integration Security Framework Using Honeypots," *Mobile Information Systems*, vol. 2022, 2022, doi: 10.1155/2022/2600712.
- [35] M. Rahouti, K. Xiong, N. Ghani, and F. Shaikh, "SYNGuard: Dynamic threshold-based SYN flood attack detection and mitigation in software-defined networks," *IET Networks*, vol. 10, no. 2, 2021, doi: 10.1049/ntw2.12009.
- [36] S. R. M. Zeebaree, K. Jacksi, and R. R. Zebari, "Impact analysis of SYN flood DDoS attack on HAProxy and NLB cluster-based web servers," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 19, no. 1, 2020, doi: 10.11591/ijeecs.v19.i1.pp505-512.
- [37] S. R. M. Zeebaree, K. H. Sharif, and R. M. Mohammed Amin, "Application Layer Distributed Denial of Service Attacks Defense

- Techniques : A review,” Academic Journal of Nawroz University, vol. 7, no. 4, 2018, doi: 10.25007/ajnu.v7n4a279.
- [38] H. TAŞÇI, S. GÖNEN, M. A. BARIŞKAN, G. KARACAYILMAZ, B. ALHAN, and E. N. YILMAZ, “Password Attack Analysis Over Honeypot Using Machine Learning Password Attack Analysis,” Turkish Journal of Mathematics and Computer Science, vol. 13, no. 2, 2021, doi: 10.47000/tjmcs.971141.
- [39] 3Gpp-Ts-23.107, “3GPP TS 23.107: Quality of Service (QoS) Concept and Architecture,” 3GPP:Technical Specification Group Services and System Aspects., vol. 0, no. Release 1999, 2009.
- [40] F. L. Rodríguez, U. S. Dias, D. R. Campelo, R. de O. Albuquerque, S. J. Lim, and L. J. G. Villalba, “QoS management and flexible traffic detection architecture for 5G mobile networks,” Sensors (Switzerland), vol. 19, no. 6, 2019, doi: 10.3390/s19061335.
- [41] N. A. majeed Alhammadi, “Comparative study between (SVM) and (KNN) classifiers by using (PCA) to improve of intrusion detection system,” Iraqi Journal of Intelligent Computing and Informatics (IJICI), vol. 1, no. 1, 2022, doi: 10.52940/ijici.v1i1.4.

# Hybrid Local Search Algorithm for Optimization Route of Travelling Salesman Problem

Muhammad Khahfi Zuhanda<sup>1</sup>, Noriszura Ismail<sup>2</sup>, Rezzy Eko Caraka<sup>3</sup>, Rahmad Syah<sup>4</sup>, Prana Ugiana Gio<sup>5</sup>

Informatics Engineering Study Program, Faculty of Engineering, Universitas Medan Area, Medan, Indonesia<sup>1</sup>

Department of Mathematical Sciences- Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Selangor, Malaysia<sup>2</sup>

Research Organization for Electronics and Informatics-National Research and Innovation Agency (BRIN), Bandung, Indonesia<sup>3</sup>

Informatics Engineering Study Program-Faculty of Engineering, Universitas Medan Area, Medan, Indonesia<sup>4</sup>

Department of Mathematics-Universitas Sumatera Utara, Medan, Indonesia<sup>5</sup>

**Abstract**—This study explores the Traveling Salesman Problem (TSP) in Medan City, North Sumatra, Indonesia, analyzing 100 geographical locations for the shortest route determination. Four heuristic algorithms—Nearest Neighbor (NN), Repetitive Nearest Neighbor (RNN), Hybrid NN, and Hybrid RNN—are investigated using RStudio software and benchmarked against various problem instances and TSPLIB data. The results reveal that algorithm performance is contingent on problem size and complexity, with hybrid methods showing promise in producing superior solutions. Statistical analysis confirms the significance of the differences between non-hybrid and hybrid methods, emphasizing the potential for hybridization to enhance solution quality. This research advances our understanding of heuristic algorithm performance in TSP problem-solving and underscores the transformative potential of hybridization strategies in optimization.

**Keywords**—Travelling Salesman Problem; heuristic algorithms; hybridization techniques algorithm performance; route optimization

## I. INTRODUCTION

Irish and British mathematicians first introduced the Traveling Salesman Problem (TSP). They were William Rowan Hamilton and Thomas Penyngton in 1800. The Traveling Salesman Problem (TSP) involves a salesman who has to travel to several points. Each point is visited only once, and the salesperson must return to the starting point again by trying the minimum path. For every  $n$  number of points, the number of routes to be traveled is  $n!$ . This problem causes no optimal solution except to calculate every possibility. So, the discussion of TSP is growing exponentially.

In life, many TSP problems are found to solve passenger delivery and pickup problems [1]–[3], drone and truck combination trips in improving customer service [4], [5], land logistics delivery problems[6]–[9], air logistics delivery [10], [11], picking up trash cars [12], automating systems on mobile robotics [13], [14] and others. The traveling salesperson problem has become one of the most studied problems in combinatorial optimization. The search for TSP solutions offers many algorithms that are fast in their calculations and produce optimal solutions.

Many scientists have tried to solve the TSP problem. Various methods are used to obtain more optimal and faster results in the calculations [15], [16]. Zhang et al. [17]

presented a variable neighborhood discrete whale optimization algorithm for TSP. Teng and Li [18] proposed a discrete firefly algorithm combining genetic algorithms for solving TSP. Several other algorithms are offered in solving TSP, such as construction tour techniques based on the Convex-hull heuristic and Nearest Neighbor (CH-NN) [19], galaxy-based search algorithm (GbSA), and embedding new ideas called clockwise search processes and operations cluster crossover [20], optimal heuristic algorithm (2-opt) with Nearest Neighbor (NN) [21]–[24], tour construction Ant Colony Algorithm [25]–[27], and Cuckoo Search Algorithm [28].

Traveling Salesman Problem (TSP) research has witnessed significant advancements over the years, marked by the development of sophisticated heuristic algorithms and the establishing of benchmark datasets like TSPLIB. These algorithmic improvements, such as Nearest Neighbor (NN), Repetitive Nearest Neighbor (RNN), and hybrid approaches, have greatly enhanced our ability to find near-optimal solutions for TSP instances. However, several challenges and unresolved issues persist in this field. Scaling complexity remains a formidable challenge, particularly when dealing with large-scale TSP instances, where the problem's exponential nature poses computational hurdles. Furthermore, despite their efficiency, heuristic algorithms do not guarantee the global optimum, leaving room for further exploration to bridge the gap between heuristic and true optimal solutions. Additionally, dynamic TSP scenarios, where cities and distances change over time, demand adaptive heuristic approaches.

Given these challenges, the presented study focusing on TSP in Medan City, North Sumatra, Indonesia, assumes particular importance and novelty. Medan City's unique geographical layout and urban complexities present a real-world context that offers practical relevance for logistics, transportation, and urban planning. The study goes beyond theoretical analysis by rigorously evaluating four heuristic algorithms—NN, RNN, Hybrid NN, and Hybrid RNN—across various problem instances specific to Medan City. This empirical examination provides valuable insights into the algorithms' performance under the city's unique conditions, aiding decision-makers in optimizing routes efficiently. Moreover, the study introduces the novel concept of hybridization techniques within the context of TSP in Medan City. The successful application of hybrid algorithms

demonstrates their potential to yield high-quality solutions, contributing to the broader knowledge of computational optimization. In essence, this study not only addresses a complex optimization problem in a real-world setting but also pioneers innovative approaches, making it a noteworthy addition to the field of TSP research.

## II. MATHEMATICAL MODEL

In this study, the mathematical model of the TSP aims to minimize travel distances.  $x_{ij}$  is the decision variable on which vertex to traverse. The set decision variable is  $\{1,0\}$ . The decision is worth 1 if arc  $(i, j)$  is passed and 0 if arc  $(i, j)$  is not passed. The  $w_{ij}$  variable is the distance between points that are traversed using point distance calculations so that it can be written down like Eq. (1).

Parameter:

- $n$  : The number of points that the salesman needs to visit.
- $w_{ij}$  : The distance or cost between two points, calculated using the Euclidean distance formula.
- $x_{ij}$  : A binary decision variable that indicates whether the salesman travels from point  $i$  to point  $j$  ( $x_{ij} = 1$ ) or not ( $x_{ij} = 0$ ).
- $|V|$  : The number of elements in the subset.
- $(a_i, b_i)$  : The position of a point in terms of its horizontal ( $a_i$ ) and vertical ( $b_i$ ) coordinates.

Formula:

$$w_{ij} = \left( (a_i - a_j)^2 + (b_i - b_j)^2 \right)^{1/2} \quad (1)$$

In general, the mathematical model for the TSP problem can be seen in Eq. (2) to Eq. (5).

$$\min \sum_{i=1}^n \sum_{j \neq i, j=1}^n w_{ij} x_{ij} \quad (2)$$

$$\sum_{i=1, i \neq j}^n x_{ij} = 1, \quad j = 1, 2, 3, \dots, n \quad (3)$$

$$\sum_{j=1, j \neq i}^n x_{ij} = 1, \quad i = 1, 2, 3, \dots, n \quad (4)$$

$$\sum_{i \in V} \sum_{j \neq i, j \in V} x_{ij} \leq |V| - 1, \quad \forall V \subseteq \{1, 2, 3 \dots n\}, |V| \geq 2 \quad (5)$$

Eq. (2) is the objective function of TSP to minimize costs. Eq. (3) and Eq. (4) guarantee that each point is traversed exactly once and returns to the starting point of departure. Eq. (5) ensures that the number of traversed vertices is not more than or equal to the specified number of vertices minus one or can be written as  $|V| - 1$ .

## III. RNN ALGORITHM

RNN and Nearest Neighbor NN are commonly used algorithms for solving the shortest route problem on a map with many points. The NN algorithm works by starting from a random point and looking for the nearest neighbour to proceed to the next point. This process is repeated until all points are connected in a closed route. The NN algorithm is simple and fast but only sometimes produces the best solution.

The RNN algorithm is a variation of the NN algorithm that works by finding the nearest neighbour at each stage but with some modifications. This algorithm looks for the nearest

neighbours for the first point, returns to the starting point and looks for the nearest neighbours yet to be connected to the route. This algorithm is repeated until all points are connected in a closed route. The RNN algorithm is generally better at finding better solutions than the NN algorithm but requires a longer computation time.

Both are heuristic algorithms that can quickly solve the shortest route problem with many points. However, keep in mind that the solution provided by the heuristic algorithm is only sometimes optimal, especially in cases with many points or in cases with many constraints. The step-by-step construction of the RNN algorithm can be read in the following step sequence:

**Step 1:** Suppose  $V = \{v_1, v_2, v_3, \dots, v_n\}$  is the point of  $n$  locations and  $d(v_i, v_j)$  is the distance between location  $v_i$  and  $v_j$ , the notation  $v_i$  is the notation of the  $i^{th}$  site.

In this case, the search engine formulates  $n$  sub-routes and has one location in each. The set of sub-routes can be denoted as follows:

$$P_1 = \{v_i; i = 1, 2, \dots, n\} \quad (6)$$

**Step 2:** In the next step, each sub-route obtained in the previous step adds a network by finding the closest location that differs from the remaining spots. So this model can be formulated based on Eq. (7). The notation  $d(v_i, v_j)$  is the Euclidean distance between locations  $v_i$  and  $v_j$ . This distance is calculated in the formula in Eq. 1.

$$P_2 = \{v_i, v_j\}; \forall v_i \in R_1;$$

$$\min d(v_i, v_j), \forall v_j \in V; i \neq j; j = 1, 2, \dots, n \quad (7)$$

**Step 3:** In the last step, each sub-route from 2 locations in  $R_2$  is expanded by finding unvisited paths. The third sub-route built is notated in Eq. 8.

$$P_3 = \{v_i, v_j, v_k\}; \forall v_i, v_j \in R_1;$$

$$\min \{d(v_j, v_k)\}, \forall v_k \in V; j \neq k; j = 1, 2, \dots, n \quad (8)$$

where  $|P_3| = n$  is the total number of routes in the set  $R_3$ . The process of building the path is continued until every location is visited. Furthermore, when the  $n$ th step has been completed, a set of routes  $P_n$  will be obtained, where  $|P_n| = n$  and each path from  $R_n$  contains  $n$  locations. The algorithm added the initial site to each path's position  $(n + 1)$  in  $P_n$  to obtain a feasible TSP route solution. The steps are continued until the best route is obtained from a set of  $n$  possible TSP routes [29].

## IV. HEURISTICS 2-OPT

2-Opt is a straightforward local search method. Croes first introduced this method to solve the TSP in 1958. However, this algorithm is used to improve the shortcomings of other algorithms, which require a long computation time. 2-Opt heuristics is a heuristic technique to solve the shortest route problem by improving existing routes. This technique is named "2-Opt" because it considers two points on a route and tries to swap the path connecting those two points with an alternative path that may be more efficient. The 2-Opt

algorithm takes an initial route consisting of several points and then considers all possible combinations of two points on the route. After that, the algorithm tries to rotate the part of the route between the two points to produce a new route. Then, the algorithm compares the total initial route distance with the resulting new total route distance and chooses the route that has the shortest total distance. Fig. 1 is an illustration of the 2-Opt algorithm.

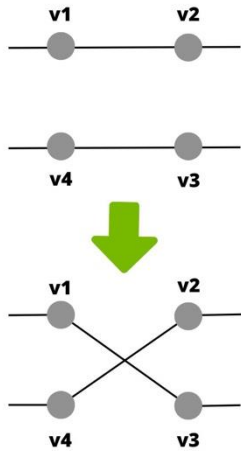


Fig. 1. An illustration of how the 2-Opt algorithm works.

In Fig. 2, there are four location points symbolized by  $v_1, v_2, v_3,$  and  $v_4$ . Initially, some edges intersect, namely  $v_1$  to  $v_2$  and  $v_3$  to  $v_4$ . This algorithm checks the distance between other adjacent points without adding new edges. The other closest sides to be checked are  $(v_1, v_3)$  and  $d(v_2, v_4)$ . The criteria for selecting the ideal side can be calculated by comparing the Euclidean distance-like Eq. (9). If the conditions in Eq. 9 are met, then the two sides  $(v_1, v_2)$  and  $(v_3, v_4)$  will be replaced with two new sides  $(v_1, v_3)$  and  $d(v_2, v_4)$ .

$$d(v_1, v_3) + d(v_2, v_4) < d(v_1, v_2) + d(v_3, v_4) \quad (9)$$

## V. HYBRID HEURISTICS

Hybrid NN is a heuristic algorithm which is a combination of two heuristic techniques, namely NN and 2-Opt. Hybrid RNN is a heuristic algorithm which is a combination of two heuristic techniques, namely RNN and 2-Opt. This algorithm is designed to solve the shortest route problem by using the advantages of both heuristic techniques. In the Hybrid Nearest Neighbor-2Opt approach, you will generate an initial solution using the NN method. Then, you'll refine this solution using the 2-opt method. This combination can be effective because the NN method provides a good starting point for the 2-opt method, which can refine the solution. The hybrid RNN-2-Opt algorithm begins by building an initial route using the RNN algorithm. After the initial route is built, the 2-Opt technique is applied to improve the route by finding a point on the route that can be exchanged so that the total route distance becomes shorter. This process is repeated until there are no more points that can be exchanged to shorten the route distance.

```
function RNN-2Opt(cities):
    best_tour = None
    best_distance = infinity
    for each city in cities:
        current_tour = NearestNeighborTour(starting from city)
        improved_tour = True
        while improved_tour:
            improved_tour = False
            for i=1 to number of cities in the tour - 1:
                for j=i+1 to number of cities in the tour:
                    new_tour = 2OptSwap(current_tour, i, j)
                    if distance of new_tour < distance of current_tour:
                        current_tour = new_tour
                        improved_tour = True
            if distance of current_tour < best_distance:
                best_tour = current_tour
                best_distance = distance of current_tour
    return best_tour
function NearestNeighborTour(starting city):
    create an empty tour
    add the starting city to the tour
    while there are unvisited cities:
        find the nearest unvisited city to the last city in the tour
        add the nearest city to the tour
    return the tour
function 2OptSwap(tour, i, j):
    return the tour up to i-1, followed by the section from i to j reversed, followed by the rest of the tour
```

Fig. 2. Pseudocode hybrid RNN.

This pseudocode follows the Hybrid RNN algorithm as follows [29]:

- 1) It starts from each city and generates a tour using the Nearest Neighbor (NN) heuristic.
- 2) Then it tries to improve this tour using the 2-Opt heuristic, making 2-Opt swaps as long as they improve the tour.
- 3) It keeps track of the best tour found so far, and once all cities have been used as starting points, it returns the best tour found.
- 4) The NearestNeighborTour function implements the Nearest Neighbor heuristic: starting from a given city, it repeatedly visits the nearest unvisited city.

## VI. EXPERIMENTAL RESULTS

In the discussion in this study, we present 100 locations in Medan city, North Sumatra, Indonesia. The locations will be analyzed for the shortest route based on location coordinates. The location to be calculated can be seen in Fig. 3. Customer location according to geographical coordinates can be seen in Fig. 3 (a), and Fig. 3(b) is the location transformation to the Cartesian point plane. This transformation makes it easy to illustrate Euclidean distance calculations.

There are four solutions proposed to be observed, namely NN, RNN, hybrid NN, and hybrid RNN. This study uses the Rstudio software with the TSP package and tspmeta for calculating heuristic solutions. A comparison of the total mileage obtained can be seen in Table I. In Table I, RNN and hybrid RNN methods have the best solution in the case of a combination of 100 location points in Medan city, North Sumatra Province, Indonesia. The table provides information about the heuristics used to solve the shortest route problem at

100 different points and the total distance travelled in kilometres. Four types of heuristics are used: NN, RNN, hybrid NN, and hybrid RNN. NN and RNN have different total distances travelled, where RNN produces a shorter total distance than NN (385.83 km compared to 435.91 km). The hybrid NN also produces a shorter total distance than the NN (414.09 km compared to 435.91 km), although it is still longer than the RNN. RNN-2Opt has the same total distance as RNN (385.83 km), so RNN-2Opt is the most effective heuristic in solving the shortest route problem at those 100 points. The results of data processing with Rstudio are visualized in Fig. 4. Fig. 4 is a travel route using the local search heuristic algorithm: NN, RNN, hybrid NN-2Opt, and hybrid NN-2Opt.

In the next stage, we tested several location points totaling 25, 40, 50, 75, 100, 150, 200, and 300 randomly generated around Medan city, North Sumatra Province. From the simulation results given, the total distance traveled in kilometers can be seen in Table II. The Table II has six problem instances, namely n25mdn, n40mdn, n50mdn, n75mdn, n100mdn, and n150mdn.

The total distance travelled by the four different algorithms is given for each problem instance. From the table, the algorithm's performance varies depending on the size of the problem instance. In the n25mdn problem instance, the hybrid NN algorithm produces the shortest total distance (202.28 km) compared to the other three algorithms. However, in the case of the n50mdn problem, the hybrid NN algorithm is no longer the best choice, while the hybrid RNN algorithm produces the shortest total distance (309.25 km). In more significant problem instances such as n100mdn, RNN and hybrid RNN beat the other two algorithms, producing the same total distance (385.83 km). At the same time, hybrid NN could only produce a total distance of 414.10 km, and NN produced a longer total distance of 435.92 km. However, remember that

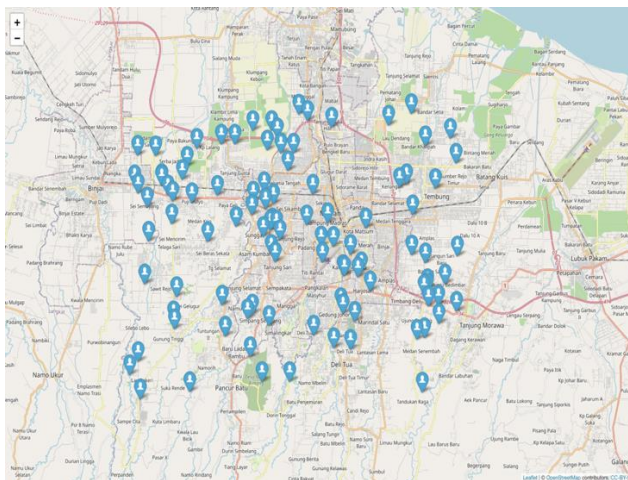
the algorithm's performance depends on the particular problem instance, so choosing an algorithm based on the characteristics and size of the problem instance to be solved is better. To find out the capabilities of this method to quickly calculate the minimum mileage. So we tested the problem with some data from TSPLIB and tried to compare the best-known solutions with our proposed method. The results of the simulation carried out can be seen in Table III.

TABLE I. TOUR LENGTH

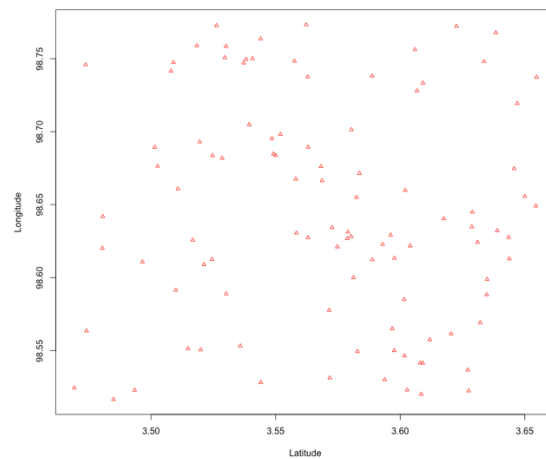
Heuristics	n	Total Distance (km)
NN	100	435.91
RNN	100	385.83
hybrid NN	100	414.09
hybrid RNN	100	385.83

From the Table III, it can be seen that the performance of the algorithm varies depending on the particular problem instance. In Berlin52, the Hybrid NN and RNN Algorithm, algorithms have the best performance, producing a total value of the shortest distance of 8182.19. However, in Ch150, the Hybrid RNN algorithm produced the best total shortest distance, 6695.24, while the NN Algorithm produced the longest total distance (8025.45). In larger problem instances such as pr299, the Hybrid NN and RNN Algorithm, algorithms can beat the other two algorithms, where both produce a better total distance value compared to the NN Algorithm and Hybrid RNN. However, these methods still need to be revised to the best-known solutions. To see a comparison of the methods is presented in Fig. 5. In Fig. 5, it can be seen that the error weight is calculated by Eq. (10). The hybrid method can reduce the error rate.

$$Error (\%) = \frac{Proposed\ Method - Best\ Known}{Best\ Known} \times 100\% \quad (10)$$



(a)



(b)

Fig. 3. One hundred location points that must be visited in Medan city, North Sumatra Province.

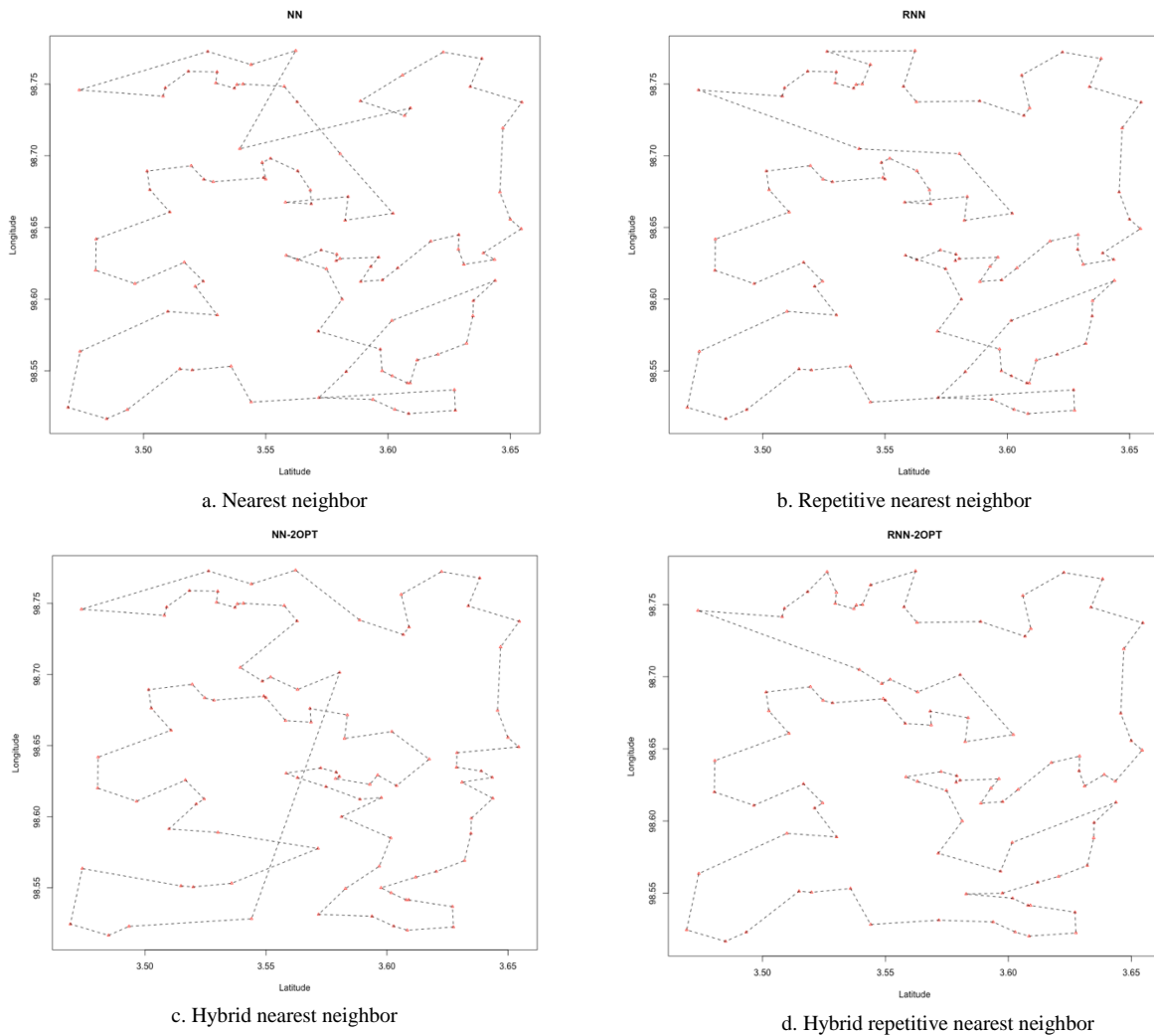


Fig. 4. Comparison of travel routes using the NN, RNN, hybrid NN-2Opt, and hybrid NN-2Opt algorithms.

TABLE II. THE DISTANCE TRAVELED WITH FOUR HEURISTIC METHODS

Problem	n	NN	Hybrid NN	RNN	Hybrid RNN
n25mdn	25	258.74	202.28	216.79	216.05
n40mdn	40	303.03	251.06	258.68	235.44
n50mdn	50	341.50	286.67	341.50	309.25
n75mdn	75	447.97	378.39	406.00	361.03
n100mdn	100	435.92	414.10	385.83	385.83
n150mdn	150	544.63	548.17	535.47	488.00
n200mdn	200	637.70	599.51	628.23	565.69
n300mdn	300	844.22	691.48	767.99	692.26

TABLE III. SIMULATION RESULTS FOR CALCULATING TOUR LENGTH USING TSPLIB DATA

Problem Name	Num. of City	Best Known	NN	Hybrid NN	RNN	Hybrid RNN
Berlin52	52	7542	8182.19	7713.03	8182.19	8182.19
Ch150	150	6528	8025.45	7656.96	7078.44	6695.24
pr299	299	48191	57901.3	49555.9	56199.22	50566.20
pr264	264	49135	54491.5	52084	54124.53	53431.19
pcb442	442	50778	58953	53044.4	59947.47	53898.70
bier127	127	118282	133971	122072	127708.80	125030.50

Fig. 5 states that the percentage errors of the NN, RNN, Hybrid NN, and Hybrid RNN algorithms in solving some shortest route problems using several datasets. The percentage error indicates how far the results produced by the algorithm are from the known best (best-known solution) for each problem. The lower the error percentage, the better the results produced by the algorithm. Based on the Fig. 5, it can be seen that hybrid NN and hybrid RNN tend to give a lower error percentage than NN and RNN. Hybrid NN even produces the lowest error value on the Berlin52 dataset, and hybrid RNN produces the lowest error value on the Ch150 dataset. This shows that using hybridization techniques can improve the quality of the solutions produced by the algorithm.

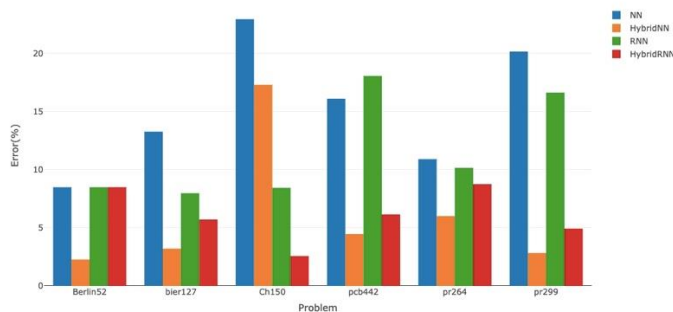


Fig. 5. Comparison of the heuristic method errors offered.

In aggregate, there is an average difference between non-hybrid and hybrid methods. The simulation results using RStudio using the `t.test()` function obtained a value of  $t = 2.8056$ ,  $df = 27$ , and  $p - value = 0.0092$  at a 95% confidence level. The t-value indicates the magnitude of the difference between the means of the two groups, in this case, the non-hybrid and hybrid methods. A larger t-value indicates a greater difference between the means. The degrees of freedom (df) represent the independent observations to estimate a parameter. In this case, it indicates the number of observations minus the number of parameters estimated, which is equal to 27. The p-value is the probability of obtaining a t-value as extreme as or more extreme than the observed t-value, assuming the null hypothesis is true. In this case, the null hypothesis is that there is no significant difference between the means of the non-hybrid and hybrid methods. A p-value of 0.0092 suggests that the probability of obtaining such an extreme t-value under the null hypothesis is less than 0.01. Therefore, the result is statistically significant at a 95% confidence level. In summary, the simulation results

suggest a significant average difference between non-hybrid and hybrid methods in terms of their performance, with the hybrid methods outperforming the non-hybrid methods on average.

## VII. DISCUSSION

In this study, the authors delve into the challenging Traveling Salesman Problem (TSP) using a variety of heuristic algorithms. Their research focuses on 100 locations in Medan city, North Sumatra, Indonesia, and aims to find the shortest route based on geographical coordinates. We transform geographical coordinates into Cartesian points to facilitate this analysis, making Euclidean distance calculations more accessible. Four heuristic solutions are proposed for examination: Nearest Neighbor (NN), Repetitive Nearest Neighbor (RNN), Hybrid NN, and Hybrid RNN. RStudio software, coupled with the TSP package and `tspmeta`, is employed for heuristic solution calculations. The study sheds light on the performance of these algorithms, particularly focusing on the total mileage obtained, which is compared across different problem instances and known benchmarks.

The results of this study reveal that the performance of heuristic algorithms is contingent upon the specific problem instance under consideration. In smaller problem instances, such as `n25mdn`, the Hybrid NN algorithm proves to be the most efficient in producing the shortest total distance. Nevertheless, as problem instances grow in complexity, like `n50mdn` and `n100mdn`, the RNN and Hybrid RNN consistently outperform the other algorithms, yielding identical total distances in the case of `n100mdn`. The choice of the algorithm should be tailored to the problem's characteristics and scale, highlighting the nuanced nature of TSP problem-solving.

The study evaluates the proposed heuristic methods with known best solutions from the TSPLIB data. This comparative analysis offers valuable insights into algorithmic performance under standardized benchmarks. For instance, the Berlin52 dataset demonstrates that Hybrid NN and RNN algorithms attain the best results, aligning with the best-known solution. Conversely, in the Ch150 dataset, the Hybrid RNN algorithm emerges as the frontrunner, boasting the best total shortest distance. This highlights the potential of hybridization techniques, like combining NN and RNN with 2-Opt, to yield superior-quality solutions and reduce error rates vis-à-vis non-hybrid methods.



The paper augments its findings with statistical rigour, employing a t-test to substantiate the significance of differences between non-hybrid and hybrid methods. The results confirm that, on average, hybrid methods surpass their non-hybrid counterparts, underscoring their potential for achieving more optimal solutions.

This study carries substantial implications for addressing TSP and akin optimization challenges. Algorithm selection should be a nuanced process tailored to the problem's size and intricacies. The fusion of heuristic algorithms, as demonstrated in hybridization techniques, holds promise for elevating solution quality. Benchmarking against established datasets like TSPLIB serves as a litmus test for algorithmic reliability. Importantly, while heuristic methods may not always secure the global optimum, their swiftness renders them indispensable for tackling real-world routing and scheduling problems. This research advances the understanding of heuristic algorithm performance within the TSP domain and underscores the transformative potential of hybridization strategies in optimization problem-solving.

### VIII. CONCLUSION

TSP is a complex combinatorial problem in calculations to find the best combination. Because it is challenging to calculate the final solution with the shortest tour length globally, however, the heuristic approach method can be used. This heuristic method is a method that can calculate a solution quickly with a pretty good solution. However, it is still inferior to modern and exact metaheuristic methods in finding solutions, but the speed of finding solutions cannot be doubted because they can calculate quickly. Therefore this method is still widely used in various applications in the real world. The results of this study indicate that the offered hybrid method can correct errors from the usual heuristic methods. The combination of NN and RNN algorithms with 2-Opt provides a better solution. The hybrid method minimized the percentage of errors compared to the non-hybrid method. The 2-Opt technique can be used in combination with other heuristic algorithms such as NN or RNN to improve the quality of the resulting solutions. However, although the 2-Opt technique can increase the efficiency and quality of solutions, this technique does not guarantee that the given solution will always be optimal.

### REFERENCES

[1] J. G. Lopes Filho, M. C. Goldberg, E. F. Gouvea Goldberg, and V. A. Petch, "Traveling salesman problem with optional bonus collection, pickup time and passengers," *Revista de Informatica Teorica e Aplicada*, vol. 27, no. 1, 2020, doi: 10.22456/2175-2745.93733.

[2] Z. Lyu, D. Pons, J. Chen, and Y. Zhang, "Developing a Stochastic Two-Tier Architecture for Modelling Last-Mile Delivery and Implementing in Discrete-Event Simulation," *Systems*, vol. 10, no. 6, 2022, doi: 10.3390/systems10060214.

[3] Z. Zhang, H. Liu, M. C. Zhou, and J. Wang, "Solving Dynamic Traveling Salesman Problems With Deep Reinforcement Learning," *IEEE Trans Neural Netw Learn Syst*, 2021, doi: 10.1109/TNNLS.2021.3105905.

[4] R. Roberti and M. Ruthmair, "Exact methods for the traveling salesman problem with drone," *Transportation Science*, vol. 55, no. 2, 2021, doi: 10.1287/TRSC.2020.1017.

[5] M. Dell'Amico, R. Montemanni, and S. Novellani, "Matheuristic algorithms for the parallel drone scheduling traveling salesman

problem," *Ann Oper Res*, vol. 289, no. 2, 2020, doi: 10.1007/s10479-020-03562-3.

[6] P. Baniasadi, M. Foumani, K. Smith-Miles, and V. Ejev, "A transformation technique for the clustered generalized traveling salesman problem with applications to logistics," *Eur J Oper Res*, vol. 285, no. 2, 2020, doi: 10.1016/j.ejor.2020.01.053.

[7] M. K. Zuhanda, S. Suwilo, O. S. Sitompul, and M. Mardinarsih, "A combination k-means clustering and 2-opt algorithm for solving the two echelon e-commerce logistic distribution," *Logforum*, vol. 18, no. 2, pp. 213–225, Jun. 2022, doi: 10.17270/J.LOG.2022.734.

[8] M. K. Zuhanda *et al.*, "Optimization of Vehicle Routing Problem in the Context of E-commerce Logistics Distribution," Mar. 2023.

[9] M. K. Zuhanda *et al.*, "Supply chain strategy during the COVID-19 terms: sentiment analysis and knowledge discovery through text mining," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 30, no. 2, p. 1120, May 2023, doi: 10.11591/ijeecs.v30.i2.pp1120-1127.

[10] Muren, J. Wu, L. Zhou, Z. Du, and Y. Lv, "Mixed steepest descent algorithm for the traveling salesman problem and application in air logistics," *Transp Res E Logist Transp Rev*, vol. 126, 2019, doi: 10.1016/j.tre.2019.04.004.

[11] F. Kong, J. Li, B. Jiang, H. Wang, and H. Song, "Trajectory Optimization for Drone Logistics Delivery via Attention-Based Pointer Network," *IEEE Transactions on Intelligent Transportation Systems*, 2022, doi: 10.1109/TITS.2022.3168987.

[12] C. X. Lou, J. Shuai, L. Luo, and H. Li, "Optimal transportation planning of classified domestic garbage based on map distance," *J Environ Manage*, vol. 254, 2020, doi: 10.1016/j.jenvman.2019.109781.

[13] K. Hernandez, B. Bacca, and B. Posso, "Multi-goal path planning autonomous system for picking up and delivery tasks in mobile robotics," *IEEE Latin America Transactions*, vol. 15, no. 2, 2017, doi: 10.1109/TLA.2017.7854617.

[14] S. Piao, Z. Ba, L. Su, D. Koutsonikolas, S. Li, and K. Ren, "Automating CSI Measurement with UAVs: From Problem Formulation to Energy-Optimal Solution," in *Proceedings - IEEE INFOCOM*, 2019, doi: 10.1109/INFOCOM.2019.8737613.

[15] A. P. U. Siahaan *et al.*, "Comparative study of prim and genetic algorithms in minimum spanning tree and Travelling Salesman Problem," *International Journal of Engineering and Technology(UAE)*, vol. 7, no. 4, 2018, doi: 10.14419/ijet.v7i4.20606.

[16] M. K. Zuhanda, H. Mawengkang, S. Suwilo, Mardinarsih, and O. S. Sitompul, "Logistics distribution supply chain optimization model with VRP in the context of E-commerce," in *AIP Conference Proceedings*, 2023, doi: 10.1063/5.0128465.

[17] J. Zhang, L. Hong, and Q. Liu, "An improved whale optimization algorithm for the traveling salesman problem," *Symmetry (Basel)*, vol. 13, no. 1, 2021, doi: 10.3390/sym13010048.

[18] L. Teng and H. Li, "Modified discrete Firefly Algorithm combining genetic algorithm for traveling salesman problem," *Telkommika (Telecommunication Computing Electronics and Control)*, vol. 16, no. 1, 2018, doi: 10.12928/TELKOMNIKA.V16I1.4752.

[19] E. O. Asani, A. E. Okeyinka, and A. A. Adebisi, "A Construction Tour Technique for Solving the Travelling Salesman Problem Based on Convex Hull and Nearest Neighbour Heuristics," in *2020 International Conference in Mathematics, Computer Engineering and Computer Science, ICMCECS 2020*, 2020, doi: 10.1109/ICMCECS47690.2020.240847.

[20] A. P. Ang and D. Jitkongchuen, "The cluster crossover operation for the symmetric Travelling Salesman Problem," *ECTI Transactions on Computer and Information Technology*, vol. 12, no. 2, 2018, doi: 10.37936/ecti-cit.2018122.132018.

[21] D. R. Singh, M. K. Singh, and T. Singh, "A hybrid heuristic algorithm for the Euclidean traveling salesman problem," in *International Conference on Computing, Communication and Automation, ICCCA 2015*, 2015, doi: 10.1109/CCAA.2015.7148514.

[22] S. Klootwijk, B. Manthey, and S. K. Visser, "Probabilistic analysis of optimization problems on generalized random shortest path metrics," *Theor Comput Sci*, vol. 866, 2021, doi: 10.1016/j.tcs.2021.03.016.

- [23] V. Ilin, D. Simić, S. D. Simić, and S. Simić, "Hybrid Genetic Algorithms and Tour Construction and Improvement Algorithms Used for Optimizing the Traveling Salesman Problem," in *Advances in Intelligent Systems and Computing*, 2021. doi: 10.1007/978-3-030-57802-2\_51.
- [24] A. Agrawal, N. Ghune, S. Prakash, and M. Ramteke, "Evolutionary algorithm hybridized with local search and intelligent seeding for solving multi-objective Euclidian TSP," *Expert Syst Appl*, vol. 181, 2021, doi: 10.1016/j.eswa.2021.115192.
- [25] M. Mavrovouniotis, S. Yang, M. Van, C. Li, and M. Polycarpou, "Ant colony optimization algorithms for dynamic optimization: A case study of the dynamic travelling salesperson problem [Research Frontier]," *IEEE Comput Intell Mag*, vol. 15, no. 1, 2020, doi: 10.1109/MCI.2019.2954644.
- [26] M. Mavrovouniotis, F. M. Muller, and S. Yang, "Ant Colony Optimization with Local Search for Dynamic Traveling Salesman Problems," *IEEE Trans Cybern*, vol. 47, no. 7, 2017, doi: 10.1109/TCYB.2016.2556742.
- [27] M. Mavrovouniotis, F. M. Müller, and S. Yang, "An ant colony optimization based memetic algorithm for the dynamic Travelling Salesman Problem," in *GECCO 2015 - Proceedings of the 2015 Genetic and Evolutionary Computation Conference*, 2015. doi: 10.1145/2739480.2754651.
- [28] V. Bhavana, V. Ramesh, and M. Sivagami, "Implementing discrete cuckoo search algorithm for TSP using MPI and beowulf cluster," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 8, 2019.
- [29] Md. A. Rahman and H. Parvez, "Repetitive Nearest Neighbor Based Simulated Annealing Search Optimization Algorithm for Traveling Salesman Problem," *OALib*, vol. 08, no. 06, 2021, doi: 10.4236/oalib.1107520.

# A Systematic Literature Review of Computational Studies in Aquaponic System

## Literature Review of Computational Studies in Aquaponic System

Khaoula Taji<sup>1</sup>, Ali Sohail<sup>2</sup>, Yassine Taleb Ahmad<sup>3</sup>, Ilyas Ghanimi<sup>4</sup>, Sheeba Ilyas<sup>5</sup>, Fadoua Ghanimi<sup>6</sup>

Electronic Systems-Information Processing-Mechanics and Energy laboratory-IBN Tofail University Kenitra Faculty of Sciences, Kenitra, Morocco<sup>1,4,6</sup>

Department of Computer Science, Minhaj University Lahore, Lahore, Pakistan<sup>2,5</sup>

Engineering Science Laboratory, Ibn Tofail University, ENSA, Kenitra, Morocco<sup>3</sup>

**Abstract**—The word aquaponics means the growth of aquatic organisms as well as plants in the controlled environment. As the nutrients used for sustainable plant growth is obtained from aquatic organisms and the nutrients that are absorbed by the plants remediate the water for the aquatic life. The advancement in the computational studies plays a vital role in every field of life. The aim of the proposed study is to deeply analyze the computational studies that used IoT, AI, Machine learning and deep learning for aquaponic systems between the years 2019 to 2022. The literature survey deeply discuss the proposed methodology, comprehends the fundamental researches, tool, advantages, limitations, concepts, and results of the recent studies proposed by the researchers in context of aquaponic system. The proposed study extract 41 research articles from these libraries based on year of publication, title, methodology, citation, paper quality and abstract. These articles are collected from seven different research article libraries including Google Scholar, Worldwide Science, IEEE Xplore, Google Books, Refseek, ACM digital Library and Science Direct. This study develops a state of the art research for the next researchers to work on the loopholes of the previous researches in an efficient manner. The results of the proposed study shows that the implementation of IoT based machine learning and deep learning framework shows state of the art results for the nutrients regulation, sensing, monitoring and controlling of the aquaponic environment. It is concluded from the proposed study that there need to be develop ensemble learning model with an efficient dataset in context of aquaponic environment.

**Keywords**—Aquaponics; machine learning; internet of thing (IoT); message queue telemetry transport; sensors; SMART aquaculture

### I. INTRODUCTION

The growth of the human population is expected to cross 10 billion till 2062. This increase in the population will create the challenge of energy, food and water for human. Computational field has a huge impact in every field of life [1] [2]. Agriculture is suffering from many problems including consumption of water, lack of land, lack of workforce etc. It is the most water consuming field that uses about 70% of water in different contexts. The word aquaponic refers to two words “Aquaculture” means to grow the water culture organisms in controlled environment and “Ponics” means growth of soil less media. The word aquaponics means the growth of aquatic organisms as well as plants in the controlled environment [3].

It is a combined production system of plant and aquatic animals in which most of the nutrient used for the development are obtained from each other. As the nutrients used for sustainable plant growth is obtained from aquatic organisms and the nutrients that are absorbed by the plants remediate the water for the aquatic life. This technology was firstly developed by the scientist in United States of America in early 1970's. The aim of this process is to create a sustainable and economic environment by efficient usage of water and nutrients, increase the profitability, Farm diversification, use of wastage, lowered the environmental impact and increasing the production to agricultural fish and plant production [4] [5] [6]. This is one of the main methods to solve the food and environmental crisis of the nature adopted by many countries worldwide. The Nutrient flow in the aquaponic system is measured constantly for the regulating the growth of the plants as the nutrients generated by the fishes is not sufficient for the growth of the plants. There are hundreds of studies proposed by the researchers for the identification and regulation of the flow of the nutrients in the aquaponic system. Fig. 1 describes the all-time interaction of the aquaponic system with computational industry [7].

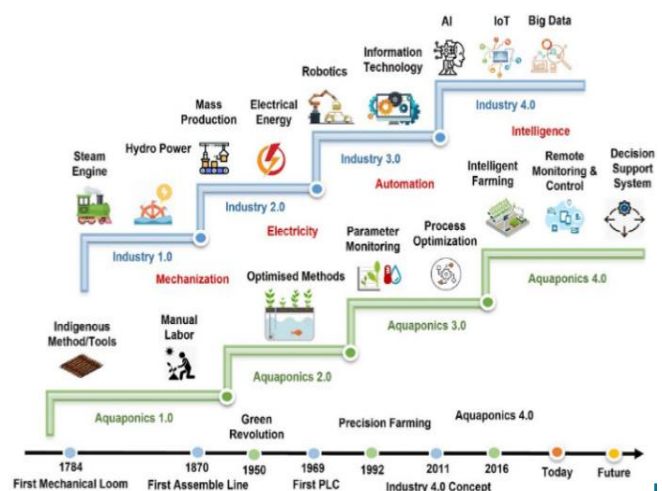


Fig. 1. Industrial revolution in the aquaponic system.

To find the best design, implementation of that design with management and maintenance is the key factor for the success of an aquaponic environment [8]. Many researcher in past

focus on different computational, statistical and mathematical tools and techniques to ensure the efficient working of an aquaponic system. The smart working of an aquaponic system based on different factors including the water quality, temperature, pH level, air flow, predictors, light intensity, humidity, Air quality, IoT sensors etc. The aim of the study is to develop a systematic survey on the latest researches proposed by different researchers for aquaponic environment. This study provides the deep analysis of the methodology, advantages and limits of the study that will create a benchmark for the new researchers in this field. A total of 41 quality research articles are deeply reviewed in this research.

The list of abbreviations used in the proposed study is illustrated in Table I

TABLE I. LIST OF ABBREVIATIONS

Word	Abbreviation
RFE	Recursive Feature Elimination
IoT	Internet of things
XGBoost	Extreme Gradient Boosting
PA	Precision Agriculture
SAR	Sodium Absorption Ratio
LDA	Linear Discrimination Analysis
BLOO	Bolstered leave one out
DTC	Decision Tree Classifier
WSN	Wireless Sensor Network
DCNNs	Deep convolutional neural networks
DB-SMOTE	Density Based Synthetic monitoring over sampling technique
SDC	Stochastic Gradient Descent
R-CNN	Regional Convolutional Neural Network
LR	Logistic Regression
SMR	Standard Metabolic
ADC	Analogue to digital convertor
US	Univariate Selection
FI	Feature importance
GNB	Gaussian Naïve Bayes
CPS	Cyber-Physical Systems
MQTT	Message Queue Telemetry Transport
NGSI	Next Generation Service Interface
ANFIS	Adaptive Neuro Fuzzy Inference System

## II. RESEARCH PLANNING

The aim of this study is to presents the latest IoT based machine learning and deep learning presented by different researcher in advancement of aquaponic system. The review method developed by Brereton et al. [9] is used in this research for reviewing the articles. This is one of the most widely used approaches for systematic literature review. This process includes five steps which are [10] discussed here in the below section.

The first phase of the research survey is planning a systematic model for work. The process of research planning include identification of the research questions, the include exclude criteria of selection, quality measurements,

identification of relevant studies and analysis of these studies. Fig. 2 explain the selection criteria of research for the proposed study [11].

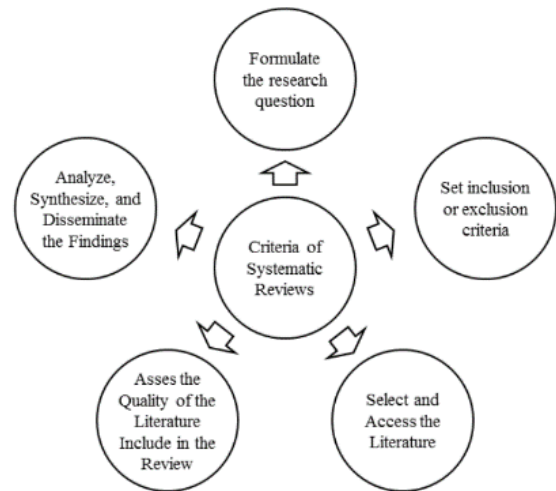


Fig. 2. Selected criteria for systematic literature review.

The first thing is to generate the research questions. The research questions for the proposed study are

RQ1: What is an Aquaponic culture?

RQ2: How the aquaponic environment works?

RQ3: What is the purpose of the proposed study?

RQ4: What are the advantages of aquaponic system?

RQ5: What are the computational studies proposed for the aquaponic system?

RQ6: What are different IoT based studies using machine learning and deep learning for aquaponics?

RQ7: What are the methodologies, advantages and loopholes of those studies?

RQ8: Comparison of the results of the previous studies?

RQ9: What are the research possibilities?

RQ10: What is the conclusion of this study?

RQ11: What are the research questions raised for the new researchers regarding aquaponic system?

The quality of the systematic literature reviews completely depend upon the selection of the related articles. In context of this, the proposed study extracts the research articles from most of the authentic and widely used article database sources. The articles are selected from seven different articles websites including Google Scholar, Worldwide Science, IEEE Xplore, Google Books, Refseek, ACM digital Library and Science Direct.

The selection is based on the quality of the research article. For searching articles from these sites different combination of words used including “Aquaculture, Aquaponics, SMART Aquaponics, IoT Based Aquaponic, Machine learning for aquaponic system, Deep learning for aquaponic system,

Nutrients regulation in Aquaponic system, Monitoring and sensing in aquaponic system, Accurate segmentation, Real time semantic, Neural network for aquaponic systems”. After applying these keywords on database libraries hundreds of articles appear. The articles are included on the basis of the relevant studies that are proposed for systematic review. This includes, excluded criteria of these articles as given below:

- Language must be English.
- Authentic Journal papers, Conference papers, Books etc.
- Papers using IoT based SMART technology
- The paper with ML and DL methods of aquaponic culture
- Papers after year 2019

The papers chosen for the proposed study are between the years 2019 and 2022. The paper published in English language and use IoT, ML, AI and DL methods are accepted for processing.

Exclude:

- Paper in language other than English
- Papers published before 2019
- Paper with method other than AI, IoT, ML and DL.

In the proposed study a total of 2,002 articles are screened from the article databases. From the selected papers 1,106 are rejected due to the publication year was before 2019. Rest of 896 articles is processed and 602 articles are rejected because of their irrelevancy of the subject. 294 selected articles are further processed and 79 out of them are selected based on IoT, ML, AI and DL methods used by the studies. 38 papers are rejected due to the low quality of paper. After the overall selection process overall 41 articles are processed for systematic literature review.

Fig. 3 explains the study flow of different phases using PRISMA [12] diagram for the article selection.

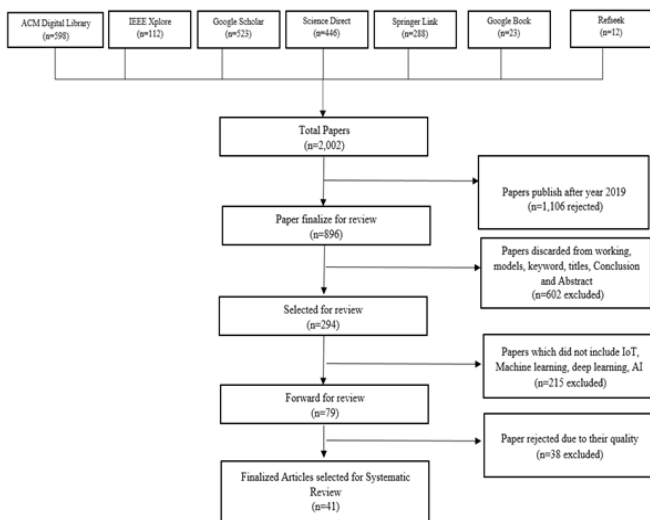


Fig. 3. Article selection criteria for systematic literature review.

### III. LITERATURE REVIEW

There are a number of computational studies developed by different researchers for the development of smart aquaponic system. In this section of the research the latest studies developed by the researchers based on IoT, Machine learning, deep learning and AI based methods between the years 2019 and 2023 are discussed. This paper deeply analyzes the advantages and disadvantages of the proposed models developed by the recent researchers.

Commercial aquaponic system helps to increase the profitability and sales revenue of aquaponic system by increasing the production. The most revenue is generated by the aquatic animals including tilapia, catfish, ornamental fish, perch, bass, trout, and bluegill [13]. The regulation of the nutrients in the aquaponic environment is one of the most discussed topic in recent years. S. B. Dhal *et al* [14] presents an IoT based system used for nutrient supply in the commercial aquaponic environment. The dataset for the study is taken from three different farms at Southeast Texas (Aquatic Greens Farm, Wolff Family Farms and Texas US Farms). This data was generated from the farms weekly over a year. From the dataset 12 predictors (Ca, Mg, Na, K, B,  $NaHCO_3$ ,  $HCO_3$ ,  $SO_4$ , Cl,  $NO_3$ ,  $NH_4$ ,  $PO_4$ ) and 211 observations are generated. On this dataset the features are extracting using pairwise correlation matrix and Recursive Feature Elimination (RFE). Machine learning algorithm XGBoost is used for generating F-score of the features and ExtraTreesClassifier is used for RFE [15]. The experiment was carried out in the cycle of 21 days. Two predictor calcium and ammonium is identified and regulate by using this system. The cost of the proposed model is decreased by 75% as compared to the existing models that are used for nutrient regulation [14].

M. A. Zamora-Izquierdo, J. Santa, J. A. Martínez, V. Martínez, and A. F. Skarmeta presents [16] IoT based on edge computing system to enhance PA. The model works on three tiers. The local plans the CPS system connects with the aquaponic crops to collect the data and perform atomic control actions. The Edge plane used for monitoring and managing the PA task and cloud plane host and record the data in FIWARE deployment. MQTT and NSGI protocols are used communicating and accessing with the cloud. The working of this system is shown in Fig. 4 [16].

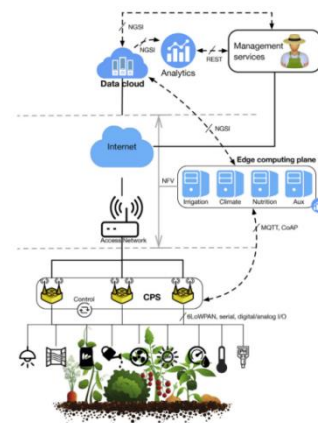


Fig. 4. Precision Agriculture based on IoT and edge computing.

The experiment is carried in Greenhouse of CEBAS-CSIC, Spain. Nutrient solution unit, irrigation unit, disinfection unit, purification unit and climate unit is used to facilitate the greenhouse environment. This research provides the water saving of more than 30% with upto 80% nutrients. This study provides the new path for the researcher to implement the PA platform in future crops. R. Barosa, S. I. S. Hassen, and L. Nagowah [17] combine hydroponic with conventional agriculture system to develop Plantabo Aevum system. The IoT devices are used for continue monitoring the environmental factor and providing the real time feedback. Live cameras are used for continue image capturing that is used for image processing. With the detection of the main features of the plant leaf the system detect the disease in it and generated the report on mobile application. The dataset is taken from the 50 leafs of four different type of plants including chilli, eggplant, mandarin and citrus. Machine learning algorithm decision tree is used for the classification. The working is implemented on OpenCV toolbox. This study detect the leafs of different species of plants accurately using machine learning system. The implementation of deep learning on the method was the loophole of that study. A researcher developed a web based monitoring system of pH, temperature and dissolved oxygen in aquaponic system [18]. Arduino microprocessor is used for measuring the environmental factors that send the measurements to local host server. Raspberry Pi is use as a network backbone of this process that transmit the information and display the live sensor data to the website on every half second [19]. In another research [20] statistical tools with machine learning algorithms are used for the nutrient regulation for the optimal growth of plants. The dataset for that study is taken from aquaponic farms of Bryan, Caldwell, and Grimes counties in Texas. The data is collected from the plant bed and fish tanks of the farms. From the collected dataset 143 observations are collected for 24 predictors out of which 11 are chemical predictors, eight are solutions, two of them measure the hardness, and other are SAR, Alkalinity and total dissolved salt. Different dimension reduction techniques [21] i.e. XGBoost and pairwise correlation matrix applied on the dataset for defining the nutrient concentration of the solution. The working of the model is explained in the Fig. 5 [20].

The predicators with less importance in the dataset are removed. Error calculation techniques including Bolstered resubstituting error estimation, BLOO error estimation and Semi-bolstered Resubstitution error estimation [22] are applied on the dataset for selecting the best methods. The results show that Semi-Bolstered Resubstitution Error estimation technique gives best result for Linear Support Vector Machine. A CNN based model using machine vision is also proposed to measure the fish length [23]. This study uses the dataset of European sea buss using camera. R-CNN gives the mIoU value of 93% for fish detection. Reduction of precision bias and increasing the precision using machine vision was the loophole of that study.

A. Taufiqurrahman, A. G. Putrada, and F. Dawani [24] proposed DT regression based model for stabilizing the water temperature for trees and fishes in an aquaponic environment. Adaptive Boosting [25] algorithm is applied with DT to avoid

the model over fitting. Swirl filter and bioball filter is installed in the fish tank for extracting the waste and nitrification of bacteria from the water. Temperature sensor is installed to detect the water temperature, water heater use to heat the water when the temperature get down, fans are used to lower the water temperature if the temperature goes high. The results of the model show that the DTR model with AdaBoost shows MSE value of 0.0045 and R-square value of 0.92.

Researcher [26] proposed a deep learning model with ResNet, SegNet18 and Inceptionv3 [27][28] for monitoring and diagnosis of the nutrient deficiency in lettuce plant of aquaponic system. The dataset used by the study consists of 3000 images that are classified into four groups. These groups are based on the images with full nutrients, Phosphate deficiency, nitrogen deficiency and potassium deficiency group. The images of the dataset are divided into training, testing and validation dataset and passed through image segmentation method for labeling the images [29]. Different features are extracted from the segmented images. These features include texture features (entropy, contrast, correlation, energy), morphological features (area, parameter) and color features are extracted. The results shows the accuracies of 98.30%, 98.90%, and 97.70% of SegNet, Inceptionv3, and ResNet18 for training set and 99.29%, 98.00% and 92.5% for validation set respectively.

S. B. Dhal, M. Bagavathiannan, U. Braga-Neto, and S. Kalafatis [30] present a comparative analysis of nutrient control in aquaponic system. The dataset of this study consists of 32 predictors taken from 201 observations. DB-SMOTE algorithm applied on the dataset for balancing the dataset values [31]. RFE with ExtraTreeClassifier shows more than 90% correlation between the predictors. M. F. Taha *et al.* [32] present machine learning model for the content detection of based on spectral data. The experiment for the study was proposed at Zhejiang University china. The working of this model is completely explained in Fig. 6.

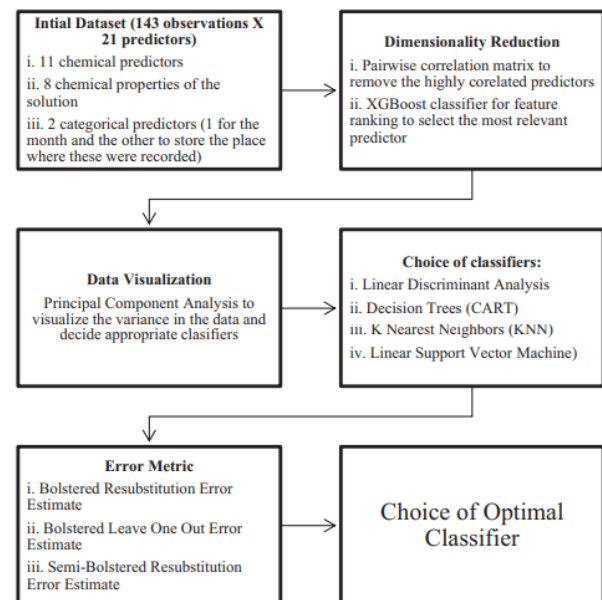


Fig. 5. The process of decision support system for nutrients regulation in aquaponic system.

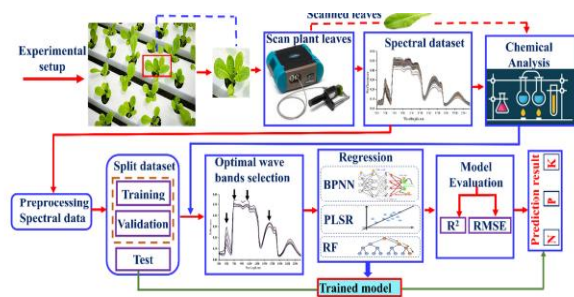


Fig. 6. The working of machine learning based Nutrient content detection model.

The spectral reflectance of the leaf [33] is measure by FieldSpec4, Pro FR portable spectroradiometer. Three machine learning models including Random Forest (RF), partial least square regression (PLSR) and backpropagation neural network (BNN) is used to classify the nutrients. The BNN algorithm shows the highest predictive accuracy of 97.2 for nitrogen content. While the highest predictive value of 0.94 and 0.96 for phosphate and potassium is obtained by RF algorithm.

In another research machine learning vision based system is employed for the lettuce growth classification. The process of lettuce development took around 45-55 days for the growth in vegetative, development and harvest cycle. The dataset of lettuce images for the study contains 300 images taken from the aquaponic system developed by Rizal, Philippines. The data extractions from these images are done by using different machine learning based feature extraction techniques in MATLAB. This classification is performed by three machine learning algorithms KNN, L-SVM and LR. The model gives the classification accuracy of 91.67% by KNN algorithm, 80% with L-SVM and 66.7% with LR algorithm [34].

Wireless technology also have a vital role in the aquaponic farming. In the latest researches the researchers presented an IoT Based system for the monitoring, regulating and controlling of the aquaponic systems [35]. T. Khaoula, R. A. Abdelouahid, I. Ezzahoui, and A. Marzak [36] presents AI and IoT based system for controlling the water quality of aquaponic system using different sensors and actuators. The system consists physical layer that consists of sensors, gateway layer that includes NodeMCU for data collection, the middleware layer that is responsible for publishing the semantics done by MQTT and the application layer use for providing the interface. There are different types of sensors including pH sensor, water level sensor, humidity sensor, temperature sensor, EC sensor, soil measure etc. used in the study for collecting the data. Haryanto, M. Ulum, A. F. Ibadillah, R. Alfita, K. Aji, and R. Rizkyandi, presents smart IoT-based system for controlling the nutrients in aquaponic system. The working of this model is explained in Fig. 7 [25].

The results of this model shows the accuracy value of 99.94% for ultrasonic sensor and 92.53% for pH sensor. In IoT based deep learning approach use edge computing for aquaponic monitoring system. The system contain four subsystems for greenhouse sensors form plant growth, aquaponic control, growth monitoring and data uploading. DHT11 sensor is used for temperature sensing, BH1750 is

used for light sensing, HC-SR04 is used for ultrasonic sensing, and SEN0161 is used for sensing pH in the current scenario. Mask-RCNN architecture is used for instance segmentation from the 250 images of fish dataset. The model shows the precision, recall and F1 score of 0.94, 0.96 and 0.95, respectively [38].

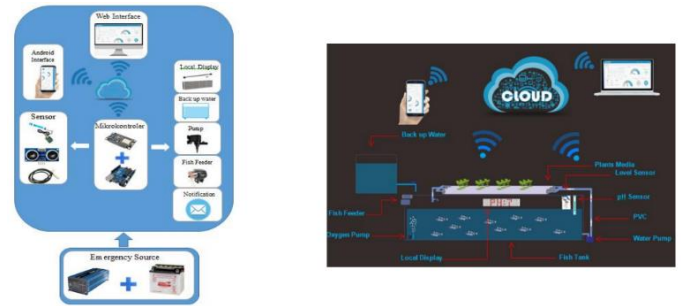


Fig. 7. System design of smart IoT based aquaponic system.

C. Lee and Y. J. [39] also present the cloud system for fish based IoT metabolism in aquaponic system through oxygen transfer rate model. IR distance sensor is used in the fish tank for finding the locomotion of the fishes in water tank. ADC is used to convert the IR signals to integer value. The dataset used in the study obtained from 27 fishes of different types in the water tank. The fish metabolic rate is used to determine the oxygen consumption level of fishes. The study shows that the pH, temperature of water and dissolved oxygen affect the metabolic activity of fishes in aquaponic system. The periodic regression of the model is carried out on ThingSpeak cloud computing platform [40]. Wang P. Mpofu, S. H. Kembo, S. Jacques, and N. Chitiyo [41] suggest IoT based household aquaponic system for food production. This was an offline-First system for overcoming the challenges of cloud computing systems in low budget household scenarios by using the Edge and Fog computing [42]. This system is deployed using LAN connection to remove the dependency of active internet connection. Edge computing is utilized to constantly check the water flow in the water pumps. This detection was based on the sound detection scenario of water in the water pumps. Raspberry Pi network is used for Fog computing. This research shows that the Edge and Fog computing system shows the cheap and efficient results in household aquaponic system.

S. C. Lauguico, R. I. S. Concepcion, J. D. Alejandrino, R. R. Tobias, and E. P. Dadios [43] classify the lettuce life in aquaponic system using machine learning for texture classification. The dataset used for the study is taken from Morong, Rizal, Philippines aquaponic farms. Haralick Texture Feature is used to extract the features from RGB images [44]. RFE, US and F1 feature extraction methods are used for extracting the features from the texture attributes. The extracted features are classified using machine learning algorithms GNB, SGD, LDA and DTC. Hold-Out validation and cross validation is applied on these algorithms for generating the results in the form of accuracy and F1 score. The best accuracy classification accuracy of 87.9% is obtained from DTC.

A. Reyes-Yanes, P. Martinez, and R. Ahmad [45] present computer vision based system to determine weight and growth rate of the fish and crops of little gem romaine lettuce the aquaponic environment. There are three basic methodology used in the model as model building for image preprocessing, image training and model training, prediction- correlation for image segmentation and parameter estimation for feature extraction. A total of 3150 instances of data is obtained from 1350 image dataset. The results of this system show the overall error of 18.7% mm for size of crop and 8.3% for weight of the fish. R. Abbasi, P. Martinez, and R. Ahmad [46] present ontology model for aquaponic grow beds. This knowledge modelling system automatically detects the required characteristics for an aquaponic crop. The AquaONT system is developed for decision making, GUI developed used inferred, and the design parameters are obtained by mathematical equations. The results of this research shows that the correct grow bed design gives the high crop yield and quality. In one of the latest research, industry 4.0 [47] method is implemented on the aquaponic environment. This method combines the latest computational studies including big data analysis, deep learning, robotics, IoT, AI and cloud computing for aquaponic environment. This research use the methontology model [48] to evaluate the AquaONT. The overall working of this whole process is shown in Fig. 8 [7].

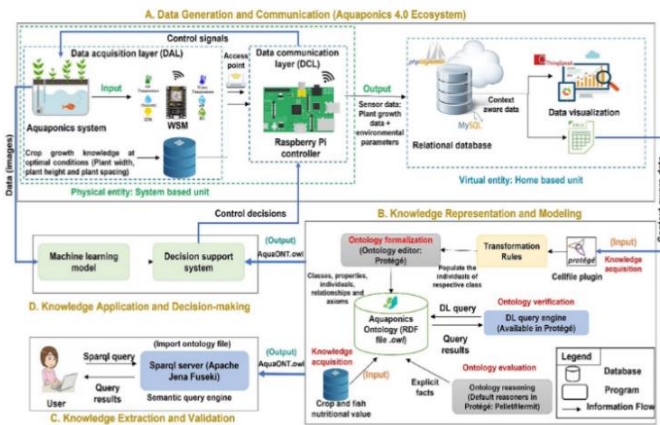


Fig. 8. An overview of AquaONT system.

This model gives the information about optimal operation of IoT devices, taking required actions on qualitative issues of fish and crops in aquaponic environment, and design configuration of grow beds based on crop characteristics while merging them with a suitable interface. These results helps the farmers to control the system of aquaponic environment in efficient manners.

A research is designed IoT based model to monitor modularization, miniaturization, and low-cost features of aquaponic system. The architecture of the model is shown in Fig. 9 [49].

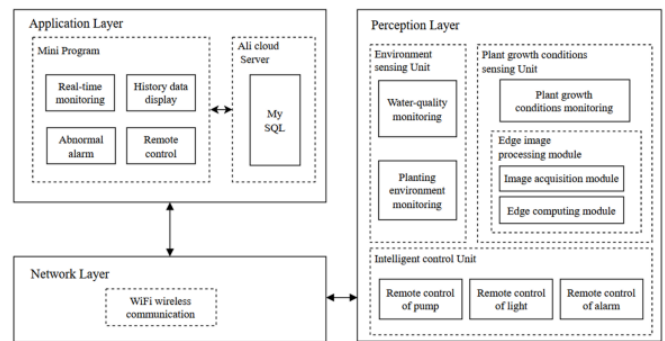


Fig. 9. Architecture of Embedded Edge computing based in IoT monitoring for aquaponic system.

The IoT based sensing consists of three layers including Application, Network and Perception layer allow the system to communicate and pass information without human interaction [50]. For this study there are three units of environment sensing, plant growth sensing, and intelligent control unit in perception layer. The model shows the environmental error rate of 5%. This system works on the edge computing using monitoring nodes, that provides the base line for the next researchers to use more nodes with strong edge sensing for the aquaponic modularization. P. Debroy and L. Seban [51] presented machine learning based study for the prediction of fruit biomass for enhancing the profit and production. The mathematical model is used to generate the dataset consists of parameters and weight of tomato in aquaponic system. Machine learning algorithm ANN and ANFSI is used for the classification. The model shows the MAE value of 0.1079 and RMSE value of 0.4582 with ANN model.

In one of the latest research AI based surrogate models [52] are implied with IoT for smart aquaponic to overcome the labor problem. This system provides the real time monitoring of water quality, temperature, pH and other parameters in aquaponic system. Fig. 10 explain the whole setup of IoT based smart pound with automatic control [53].

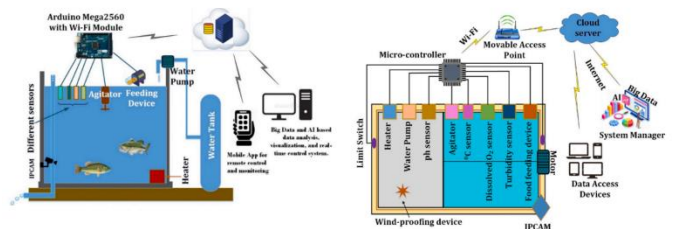


Fig. 10. IoT based smart aquaponic system.

Five sensors, five actuators along with Arduino Mega2560 are used in the system for sensing and monitoring the data. The result of the research shows  $R^2$  value of 0.94 and a MSE value of 0.0015. The comparison of latest models for the studies is explained in Table II.



TABLE II. LITERATURE REVIEW TABLE

Paper Title	Year of Publication	Methods	Dataset	Result
Smart aquaponic system based Internet of Things (IoT) [37]	2019	IoT cloud based system	Live sensors for collecting the data.	The accuracy of 99.94% for Ultrasonic sensor and 92.35% for pH sensor.
Smart Aquaponics with Disease Detection [17]	2019	IoT based system with machine learning classifier	Real time Firebase database	Detect leafs of different species for disease detection.
Smart farming IoT platform based on edge and cloud computing [16]	2019	IoT based Edge computing	Greenhouse of CEBAS-CSIC	Saving more than 30% and 80% nutrients.
Using Machine Vision to Estimate Fish Length from Images using Regional Convolutional Neural Networks [23]	2019	Machine vision using R-CNN	European sea bass image dataset	mIoU value of 93%
Lettuce life stage classification from texture attributes using machine learning estimators and feature selection processes [43]	2020	REF, US and F1 feature extraction model with GNB, LDA, DTC and SGD algorithms	Data instances from Morongo, Rizal, Philippines aquaponic farms	DTC shows the classification accuracy of 87.9%
Real-time growth rate and fresh weight estimation for little gem romaine lettuce in aquaponic grow beds [45]	2020	Computer vision based system including image processing, deep learning and regression analysis	Image dataset created by Department of environmental science of University of Alberta	The overall error of 18.7% mm for size of crop and 8.3% for weight of the fish
A Comparative Analysis of Machine Learning Algorithms Modeled from Machine Vision-Based Lettuce Growth Stage Classification in Smart Aquaponics [34]	2020	KNN, L-SVM with LR	Images dataset developed by Rizal, Philippines	Classification accuracy 91.67%
Decision Tree Regression with AdaBoost Ensemble Learning for Water Temperature Forecasting in Aquaponic Ecosystem [24]	2022	DT Regression with AdaBoost Ensemble learning	Experimental based aquaponic system at Telkom University lab	DTR model with AdaBoost shows MSE value of 0.0045 and R-square value of 0.92.
Edge Computing Based Smart Aquaponics Monitoring System Using Deep Learning in IoT Environment [38]	2020	AutoML model with gradient boost Machine learning algorithm	Images of fish dataset from Singapore Bioimaging Consortium, Singapore,	Precision, recall and F1 score of 0.94, 0.96 and 0.95 respectively.
Development of a Cloud-based IoT Monitoring System for Fish Metabolism and Activity in Aquaponics [40]	2020	Cloud based IoT monitoring system using ThingCloud Computing	Data from 27 fishes in the aquaponic environment from National Sun Yat-sen University	The pH, temperature of water and dissolved oxygen effect the metabolic activity of fishes in aquaponic system.
Utilizing a Privacy-Preserving IoT Edge and Fog Architecture in Automated Household Aquaponics [41]	2021	Edge and Fog computing based IoT system	St Peters Mbare IoT Maker space	Edge and Fog computing serve best for household aquaponic system
An Ontology model to support the automated design of aquaponic grow beds [46]	2021	Knowledge modeling approach AquaONT	LIMDA, University of Alberta	The correct grow bed design gives the high crop yield and quality.
A Machine-Learning Based IoT System for Optimizing Nutrient Supply in Commercial Aquaponic Operations [14]	2022	XGBoost and ExtraTreesClassifier With pairwise correlation matrix and Recursive Feature Elimination	Aquatic Greens Farm , Wolff Family Farms and Texas US Farms	Calcium and Ammonium predictors are identified and cost is decreased by 75%.
Nutrient optimization for plant growth in Aquaponic irrigation using Machine Learning for small training datasets [20]	2022	XGBoost and pairwise correlation matrix for dimension reduction and LDA, CART, KNN, SVM for the classification	Bryan, Caldwell, and Grimes counties	semi-Bolstered Resubstitution shows the error values of zero

Using Deep Convolutional Neural Network for Image-Based Diagnosis of Nutrient Deficiencies in Plants Grown in Aquaponics [26]	2022	Deep convolutional neural networks	3000 images of lettuce plant captured by camera (PowerShot SX720 HS)	Accuracy of 96.5%
Can Machine Learning classifiers be used to regulate nutrients using small training datasets for aquaponic irrigation? A comparative analysis [30]	2022	XGBoost and ExtraTreesClassifier	Farms in Texas	More than 90% correlation between the predictors
Using Machine Learning for Nutrient Content Detection of Aquaponics-Grown Plants Based on Spectral Data [32]	2022	Random Forest, Partial least square regression and Back propagation neural network	Spectral data self-obtained from the plant leaves.	The predictive value of $R^2p = 0.97$ for nitrogen is obtained by BNN. While RF gives $R^2p = 0.94$ for phosphate and $R^2p = 0.96$ for potassium
An ontology model to represent aquaponics 4.0 system's knowledge [7]	2022	AquaONT model with methontology approach following deep learning, computer vision and machine learning approaches	Data taken from different farms of Canada	The model show best results for optimal operation of IoT devices, qualitative issues of fish and crops, and design configuration of crop beds grow beds
Tomato Fruit Biomass Prediction Model for Aquaponics System Using Machine Learning Algorithms [51]	2022	ANN with AFNIS		MAE value of 0.1079 and RMSE value of 0.4582
A Modularized IoT Monitoring System with Edge-Computing for Aquaponics [49]	2022	Embedded Edge computing based in IoT monitoring	Real time data collection	Environmental error rate 5%.
Development of smart aquaculture farm management system using IoT and AI-based surrogate models [53]	2022	AI based surrogate model with deep CNN and IoT	Real time monitoring and collecting data	$R^2$ value of 0.94 and a MSE value of 0.0015,

IV. ANALYSIS AND DISCUSSION

Fig. 11 explain the overall working of systematic literature review for the proposed study.

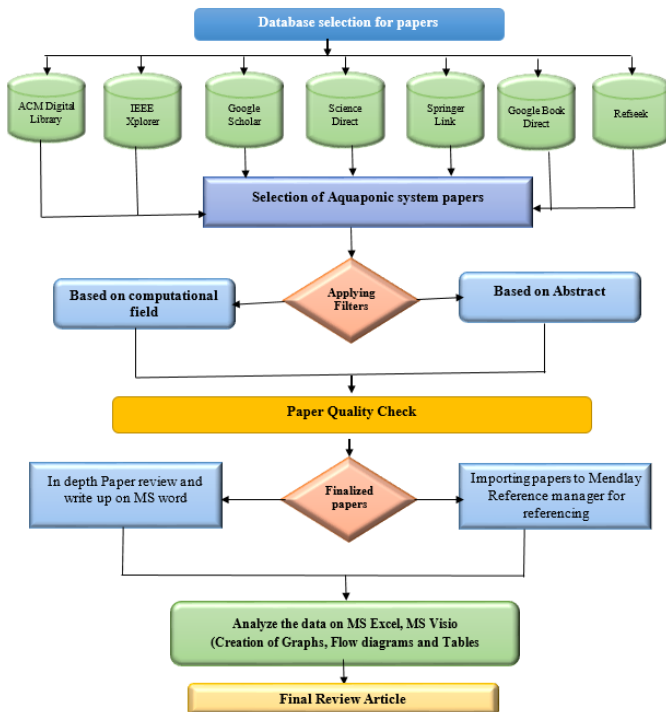


Fig. 11. Workflow of the proposed model.

In the very first step, the papers are selected from seven different known article databases. After applying different queries on these articles only 41 articles are future processed for systematic literature review. These papers are reviewed deeply as discussed in the Literature review section and the methodology, pros and cons of this study are extracted. The final draft of the study present a state of the art article for the new researchers in field of aquaponic culture to analyze the latest methods deeply and find a new direction in context of their working. All the papers that are selected are between years 2019 to 2022. The frequency distributions of the papers are shown in Fig. 12.

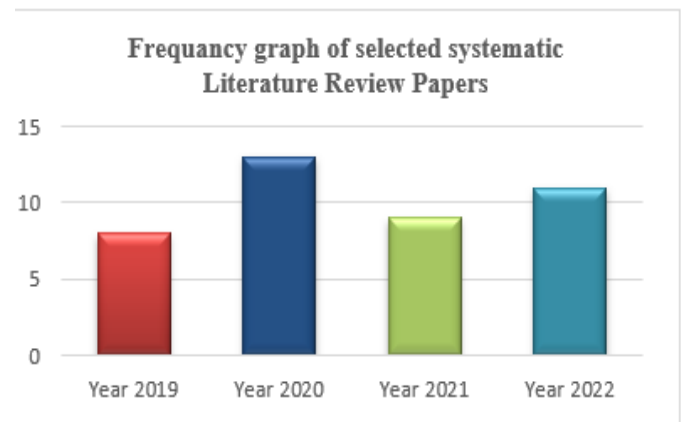


Fig. 12. Workflow graph of the proposed model.

All of the selected study gives an efficient result for aquaponic environments. Each study has its own benefits as well as drawback for the upcoming researchers. All the papers are chosen for the study are collected from high source journals and conferences with maximum citations are used in this study for maintaining the quality of the research. The heatmap diagram of the proposed study is illustrated in Fig. 13.

Figure illustrates the publication in context of year with the proposed algorithm. The selected papers are taken from different journal or conferences. The most cited papers among all along with the publishing source are explained in Table III.

In the table from 2019, it may be observed that, till date the most cited research article was IoT based edge detecting system that is published on biosystem engineering journal. Following this [23] [45] [34] has 47, 34, 31 citations respectively considered as best research articles regarding to the field.

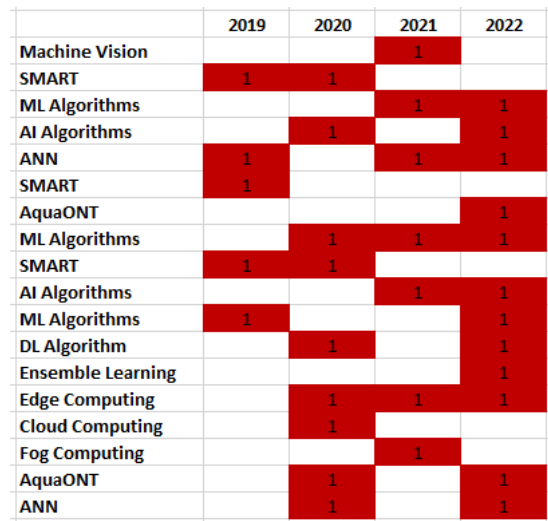


Fig. 13. Heatmap diagram of the proposed selected studies.

TABLE III. ARTICLES WITH MOST CITATIONS AND THEIR PUBLICATION JOURNALS

Author	Paper Title	Year of Publication	Methods	Journal Conference
M. A. Zamora-Izquierdo, J. Santa, J. A. Martínez, V. Martínez, and A. F. Skarmeta [16]	Smart farming IoT platform based on edge and cloud computing	2019	IoT based Edge computing	biosystems engineering
G. G. Monkman, K. Hyder, M. J. Kaiser, and F. P. Vidal [23]	Using Machine Vision to Estimate Fish Length from Images using Regional Convolutional Neural Networks	2019	Machine vision using R-CNN	Methods in Ecology and Evolution
A. Reyes-Yanes, P. Martinez, and R. Ahmad [45]	REAL-TIME GROWTH RATE AND FRESH WEIGHT ESTIMATION FOR LITTLE GEM ROMAINE LETTUCE IN AQUAPONIC GROW BEDS	2020	Computer vision based system including image processing, deep learning and regression analysis	Computer and Engineering in Agriculture
S. C. Lauguico, R. S. Concepcion, J. D. Alejandrino, R. R. Tobias, D. D. Macasaet, and E. P. Dadios [34]	A Comparative Analysis of Machine Learning Algorithms Modeled from Machine Vision-Based Lettuce Growth Stage Classification in Smart Aquaponics	2020	KNN, L-SVM with LR	International Journal of Environmental Science and Technology
C. Lee and Y. J. [40]	Development of a Cloud-based IoT Monitoring System for Fish Metabolism and Activity in Aquaponics	2020	Cloud based IoT monitoring system using ThingCloud Computing	International Journal of Environmental Science and Technology
S. C. Lauguico, R. I. S. Concepcion, J. D. Alejandrino, R. R. Tobias, and E. P. Dadios [43]	Lettuce life stage classification from texture attributes using machine learning estimators and feature selection processes	2020	REF, US and F1 feature extraction model with GNB, LDA, DTC and SGD algorithms	Methods in Ecology Evolution

### V. CONCLUSION

Aquaponic system is the growth of aquatic organisms as well as plants in the controlled environment to overcome the problem of nutrients regulation, consumption of water, lack of land, lack of workforce etc. It is one of the major discussed topics in the current scenario. This study is proposed to review the latest computational studies proposed by different researchers in aquaponic system providing the baseline for the next researchers. A total of 41 high quality research articles are choose from seven different articles database to review for the study. After deeply analyzing these researches the aim, objectives, limitations and future work of these articles are generated as illustrated in Tables II and III of the research.

It is seen in the proposed study that the most of the researchers use different IoT sensors including water quality sensor, temperature sensor, pH sensor, air flow, predictor’s sensor, light sensor, humidity sensor for constant monitoring of the aquaponic environment. There are different machine learning and deep learning models proposed by the researchers for regulating the predictors in aquaponic system.

The results of the proposed study stated that the highest sensor accuracy 99.4% for IoT ultrasonic sensor is obtained by research presented by Haryanto, M. Ulum, A. F. Ibadillah, R. Alfita, K. Aji, and R. Rizkyandi [37]. The highest classification accuracy of 96.5% is obtained by paper titled Using Deep Convolutional Neural Network for Image-Based

Diagnosis of Nutrient Deficiencies in Plants Grown in Aquaponics [26]. Paper presented by S. C. Lauguico, R. S. Concepcion, J. D. Alejandrino, R. R. Tobias, D. D. Macasaet, and E. P. Dadios [34] and S. C. Lauguico, R. I. S. Concepcion, J. D. Alejandrino, R. R. Tobias, and E. P. Dadios [43] gives the detection accuracy of 91.67% and 87.9% respectively. Smart farming IoT platform based on edge and cloud computing [12] shows the mostly cited research article use IOT based edge computing system for nutrients regulation and save upto 80% nutrients. The study [51] give the MAE value of 0.10 while [53] gives the MSE value of 0.005.

## VI. FUTURE WORK

It is seen from the proposed study that most of the researchers used real time IoT sensors for continue capturing the data and process. A few of them uses a dataset to train the ML and DL model. So to develop and use an efficient dataset in context of aquaponic environment is one of the major needs of the aquaponic study. It was also seen from the literature review that none of the research use ensemble learning method implemented on different deep learning and machine learning algorithms for aquaponic system. Deep learning algorithms with different RNN methods are not implemented by any researcher for all time in aquaponic environment.

This is the guideline for the new researcher to develop a more secure IoT based model that can use either of the ensemble learning or deep learning with RNN and CNN models to make aquaponic system more secure and sustainable.

## REFERENCES

- [1] S. Ilyas, A. A. Shah, and A. Sohail, "Order Management System for Time and Quantity Saving of Recipes Ingredients Using GPS Tracking Systems," *IEEE Access*, vol. 9, pp. 100490–100497, 2021, doi: 10.1109/ACCESS.2021.3090808.
- [2] A. A. Shah, M. K. Ehsan, A. Sohail, and S. Ilyas, "Analysis of Machine Learning techniques for identification of post translation modification in protein sequencing: A Review," in 4th International Conference on Innovative Computing, ICIC 2021, 2021, pp. 1–6. doi: 10.1109/ICIC53490.2021.9693020.
- [3] B. Yep and Y. Zheng, "Aquaponic trends and challenges – A review," *J. Clean. Prod.*, vol. 228, pp. 1586–1599, 2019, doi: 10.1016/j.jclepro.2019.04.290.
- [4] W. Lennard and S. Goddek, *Aquaponics Food Production Systems*, no. June. 2019. doi: 10.1007/978-3-030-15943-6.
- [5] S. Goddek et al., *Decoupled Aquaponics Systems*. 2019. doi: 10.1007/978-3-030-15943-6\_8.
- [6] A. J. Van Der Goot et al., "Concepts for further sustainable production of foods," *J. Food Eng.*, vol. 168, no. November 2018, pp. 42–51, 2016, doi: 10.1016/j.jfoodeng.2015.07.010.
- [7] R. Abbasi, P. Martinez, and R. Ahmad, "An ontology model to represent aquaponics 4.0 system's knowledge," *Inf. Process. Agric.*, vol. 9, no. 4, pp. 514–532, 2022, doi: 10.1016/j.inpa.2021.12.001.
- [8] M. F. Taha et al., "Recent Advances of Smart Systems and Internet of Things (IoT) for Aquaponics Automation: A Comprehensive Overview," *Chemosensors*, vol. 10, no. 8, 2022, doi: 10.3390/chemosensors10080303.
- [9] P. Brereton, B. A. Kitchenham, D. Budgen, M. Turner, and M. Khalil, "Lessons from applying the systematic literature review process within the software engineering domain," *J. Syst. Softw.*, vol. 80, no. 4, pp. 571–583, 2007, doi: 10.1016/j.jss.2006.07.009.
- [10] K. S. Khan, R. Kunz, J. Kleijnen, and G. Antes, "Five steps to conducting a systematic review," *J. R. Soc. Med.*, vol. 96, no. 3, pp. 118–121, 2003, doi: 10.1258/jrsm.96.3.118.
- [11] A. Ramdhani, M. A. Ramdhani, and A. S. Amin, "Writing a Literature Review Research Paper: A step-by-step approach," *Int. J. Basic Appl. Sci.*, vol. 03, no. 01, pp. 47–56, 2014.
- [12] L. A. Kahale et al., "PRISMA flow diagrams for living systematic reviews: a methodological survey and a proposal," *F1000Research*, vol. 10, no. March, p. 192, 2021, doi: 10.12688/f1000research.51723.1.
- [13] D. C. Love et al., "Commercial aquaponics production and profitability: Findings from an international survey," *Aquaculture*, vol. 435, pp. 67–74, 2015, doi: 10.1016/j.aquaculture.2014.09.023.
- [14] S. B. Dhal et al., "A Machine-Learning-Based IoT System for Optimizing Nutrient Supply in Commercial Aquaponic Operations," *Sensors*, vol. 22, no. 9, pp. 1–14, 2022, doi: 10.3390/s22093510.
- [15] A. Mariot, S. Sgoifo, and M. Sauli, "I gozzi endotoracici: contributo casistico-clinico (20 casi)," *Friuli Med.*, vol. 19, no. 6, 1964.
- [16] M. A. Zamora-Izquierdo, J. Santa, J. A. Martínez, V. Martínez, and A. F. Skarmeta, "Smart farming IoT platform based on edge and cloud computing," *Biosyst. Eng.*, vol. 177, no. xxxx, pp. 4–17, 2019, doi: 10.1016/j.biosystemseng.2018.10.014.
- [17] R. Barosa, S. I. S. Hassen, and L. Nagawah, "Smart Aquaponics with Disease Detection," 2nd Int. Conf. Next Gener. Comput. Appl. 2019, NextComp 2019 - Proc., pp. 1–6, 2019, doi: 10.1109/NEXTCOMP.2019.8883437.
- [18] J. P. Mandap, D. Sze, G. N. Reyes, S. Matthew Dumlaio, R. Reyes, and W. Y. Danny Chung, "Aquaponics pH Level, Temperature, and Dissolved Oxygen Monitoring and Control System Using Raspberry Pi as Network Backbone," *IEEE Reg. 10 Annu. Int. Conf. Proceedings/TENCON*, vol. 2018-October, no. 1, pp. 1381–1386, 2019, doi: 10.1109/TENCON.2018.8650469.
- [19] A. Nayyar and V. Puri, "Raspberry Pi-A Small, Powerful, Cost Effective and Efficient Form Factor Computer: A Review," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 5, no. July, pp. 720–737, 2015, [Online]. Available: <https://www.researchgate.net/publication/305668622>
- [20] S. B. Dhal, M. Bagavathiannan, U. Braga-Neto, and S. Kalafatis, "Nutrient optimization for plant growth in Aquaponic irrigation using Machine Learning for small training datasets," *Artif. Intell. Agric.*, vol. 6, pp. 68–76, 2022, doi: 10.1016/j.aiaa.2022.05.001.
- [21] K. Muhi and Z. C. Johanyák, "Dimensionality Reduction Methods Used in Machine Learning," *Műszaki Tudományos Közlemények*, vol. 13, no. 1, pp. 148–151, 2020, doi: 10.33894/mtk-2020.13.27.
- [22] U. Braga-Neto and E. Dougherty, "Bolstered error estimation," *Pattern Recognit.*, vol. 37, no. 6, pp. 1267–1281, 2004, doi: 10.1016/j.patcog.2003.08.017.
- [23] G. G. Monkman, K. Hyder, M. J. Kaiser, and F. P. Vidal, "Using machine vision to estimate fish length from images using regional convolutional neural networks," *Methods Ecol. Evol.*, vol. 10, no. 12, pp. 2045–2056, 2019, doi: 10.1111/2041-210X.13282.
- [24] A. Taufiqurrahman, A. G. Putrada, and F. Dawani, "Decision Tree Regression with AdaBoost Ensemble Learning for Water Temperature Forecasting in Aquaponic Ecosystem," 6th Int. Conf. Interact. Digit. Media, ICIDM 2020, no. Icidm, 2020, doi: 10.1109/ICIDM51048.2020.9339669.
- [25] Y.-Q. Wang, "An Analysis of the Viola-Jones Face Detection Algorithm," *Image Process. Line*, vol. 4, pp. 128–148, 2014, doi: 10.5201/ipol.2014.104.
- [26] M. F. Taha et al., "Using Deep Convolutional Neural Network for Image-Based Diagnosis of Nutrient Deficiencies in Plants Grown in Aquaponics," *Chemosensors*, vol. 10, no. 2, pp. 1–23, 2022, doi: 10.3390/chemosensors10020045.
- [27] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, 2017, doi: 10.1109/TPAMI.2016.2644615.
- [28] S. I. and M. K. E. A. Sohail, N. A. Nawaz, A. A. Shah, S. Rasheed, "A Systematic Literature Review on Machine Learning and Deep Learning Methods for Semantic Segmentation," *IEEE Access*, vol. 10, pp. 134557–134570, 2022, doi: 10.1109/ACCESS.2022.3230983.
- [29] G. Lin and W. Shen, "Research on convolutional neural network based

- on improved Relu piecewise activation function,” *Procedia Comput. Sci.*, vol. 131, pp. 977–984, 2018, doi: 10.1016/j.procs.2018.04.239.
- [30] S. B. Dhal, M. Bagavathiannan, U. Braga-Neto, and S. Kalafatis, “Can Machine Learning classifiers be used to regulate nutrients using small training datasets for aquaponic irrigation?: A comparative analysis,” *PLoS One*, vol. 17, no. 8 August, pp. 1–15, 2022, doi: 10.1371/journal.pone.0269401.
- [31] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic minority over-sampling technique,” *J. Artif. Intell. Res.*, vol. 16, no. February 2017, pp. 321–357, 2002, doi: 10.1613/jair.953.
- [32] M. F. Taha et al., “Using Machine Learning for Nutrient Content Detection of Aquaponics-Grown Plants Based on Spectral Data,” *Sustain.*, vol. 14, no. 19, 2022, doi: 10.3390/su141912318.
- [33] C. Buschmann, S. Lenk, and H. K. Lichtenthaler, “Reflectance spectra and images of green leaves with different tissue structure and chlorophyll content,” *Isr. J. Plant Sci.*, vol. 60, no. 1–2, pp. 49–64, 2012, doi: 10.1560/IJPS.60.1-2.49.
- [34] S. C. Lauguico, R. S. Concepcion, J. D. Alejandrino, R. R. Tobias, D. D. Macasaet, and E. P. Dadios, “A comparative analysis of machine learning algorithms modeled from machine vision-based lettuce growth stage classification in smart aquaponics,” *Int. J. Environ. Sci. Dev.*, vol. 11, no. 9, pp. 442–449, 2020, doi: 10.18178/ijesd.2020.11.9.1288.
- [35] A. R. Yanes, P. Martinez, and R. Ahmad, “Towards automated aquaponics: A review on monitoring, IoT, and smart systems,” *J. Clean. Prod.*, vol. 263, no. April, 2020, doi: 10.1016/j.jclepro.2020.121571.
- [36] T. Khaoula, R. A. Abdelouahid, I. Ezzahoui, and A. Marzak, “Architecture design of monitoring and controlling of IoT-based aquaponics system powered by solar energy,” *Procedia Comput. Sci.*, vol. 191, no. September, pp. 493–498, 2021, doi: 10.1016/j.procs.2021.07.063.
- [37] Haryanto, M. Ulum, A. F. Ibadillah, R. Alfita, K. Aji, and R. Rizkyandi, “Smart aquaponic system based Internet of Things (IoT),” *J. Phys. Conf. Ser.*, vol. 1211, no. 1, 2019, doi: 10.1088/1742-6596/1211/1/012047.
- [38] C. S. Arvind, R. Jyothi, K. Kaushal, G. Girish, R. Saurav, and G. Chetankumar, “Edge Computing Based Smart Aquaponics Monitoring System Using Deep Learning in IoT Environment,” 2020 IEEE Symp. Ser. Comput. Intell. SSCI 2020, no. November 2021, pp. 1485–1491, 2020, doi: 10.1109/SSCI47803.2020.9308395.
- [39] C. Lee and Y. J. Wang, “Development of a cloud-based IoT monitoring system for Fish metabolism and activity in aquaponics,” *Aquac. Eng.*, vol. 90, p. 102067, 2020, doi: 10.1016/j.aquaeng.2020.102067.
- [40] A. A. H. Mohamad, N. K. Jumaa, and S. H. Majeed, “ThingSpeak cloud computing platform based ECG diagnose system,” *Int. J. Comput. Digit. Syst.*, vol. 8, no. 1, pp. 11–18, 2019, doi: 10.12785/ijcnds/080102.
- [41] P. Mpofu, S. H. Kembo, S. Jacques, and N. Chitiyo, “Utilizing a privacy-preserving iot edge and fog architecture in automated household aquaponics,” *Proc. Int. Conf. Ind. Eng. Oper. Manag.*, vol. 59, no. November, pp. 2281–2288, 2020.
- [42] A. Sunyaev, “Fog and Edge Computing,” *Internet Comput.*, no. September, pp. 237–264, 2020, doi: 10.1007/978-3-030-34957-8\_8.
- [43] S. C. Lauguico, R. I. S. Concepcion, J. D. Alejandrino, R. R. Tobias, and E. P. Dadios, “Lettuce life stage classification from texture attributes using machine learning estimators and feature selection processes,” *Int. J. Adv. Intell. Informatics*, vol. 6, no. 2, pp. 173–184, 2020, doi: 10.26555/ijain.v6i2.466.
- [44] P. Brynolfsson et al., “Haralick texture features from apparent diffusion coefficient (ADC) MRI images depend on imaging and pre-processing parameters,” *Sci. Rep.*, vol. 7, no. 1, 2017, doi: 10.1038/s41598-017-04151-4.
- [45] A. Reyes-Yanes, P. Martinez, and R. Ahmad, “Real-time growth rate and fresh weight estimation for little gem romaine lettuce in aquaponic grow beds,” *Comput. Electron. Agric.*, vol. 179, no. September, p. 105827, 2020, doi: 10.1016/j.compag.2020.105827.
- [46] R. Abbasi, P. Martinez, and R. Ahmad, “An ontology model to support the automated design of aquaponic grow beds,” *Procedia CIRP*, vol. 100, pp. 55–60, 2021, doi: 10.1016/j.procir.2021.05.009.
- [47] L. Thames and D. Schaefer, *Industry 4.0: An Overview of Key Benefits, Technologies, and Challenges*, no. April 2019. 2017. doi: 10.1007/978-3-319-50660-9\_1.
- [48] M. Fernandez, A. Gómez-Pérez, and N. Juristo, “Methontology: from ontological art towards ontological engineering,” *Proc. AAAI97 Spring Symp. Ser. Ontol. Eng.*, no. May 2014, pp. 33–40, 1997, [Online]. Available: <http://speech.inesc.pt/~joana/prc/artigos/06c METHONTOLOGY from Ontological Art towards Ontological Engineering - Fernandez, Perez, Juristo - AAAI - 1997.pdf>
- [49] S. Wan, K. Zhao, Z. Lu, J. Li, T. Lu, and H. Wang, “A Modularized IoT Monitoring System with Edge-Computing for Aquaponics,” *Sensors*, vol. 22, no. 23, 2022, doi: 10.3390/s22239260.
- [50] P. Sethi and S. R. Sarangi, “Internet of Things: Architectures, Protocols, and Applications,” *J. Electr. Comput. Eng.*, vol. 2017, 2017, doi: 10.1155/2017/9324035.
- [51] P. Debroy and L. Seban, “A Tomato Fruit Biomass Prediction Model for Aquaponics System Using Machine Learning Algorithms,” *IFAC-PapersOnLine*, vol. 55, no. 1, pp. 709–714, 2022, doi: 10.1016/j.ifacol.2022.04.116.
- [52] M. Frangos, Y. Marzouk, K. Willcox, and B. van B. Waanders, “Surrogate and reduced-order modeling: A comparison of approaches for large-scale statistical inverse problems,” *Large-Scale Inverse Probl. Quantif. Uncertain.*, pp. 123–149, 2010, doi: 10.1002/9780470685853.ch7.
- [53] M. C. Chiu, W. M. Yan, S. A. Bhat, and N. F. Huang, “Development of smart aquaculture farm management system using IoT and AI-based surrogate models,” *J. Agric. Food Res.*, vol. 9, no. August, p. 100357, 2022, doi: 10.1016/j.jafr.2022.100357.

# Cocoa Pods Diseases Detection by MobileNet Confluence and Classification Algorithms

Diarra MAMADOU<sup>1</sup>, Kacoutchy Jean AYIKPA<sup>2</sup>, Abou Bakary BALLO<sup>3</sup>, Brou Médard KOUASSI<sup>4</sup>

LaMI, Université Felix Houphouët-Boigny, Abidjan, CÔTE D'IVOIRE<sup>1, 3, 4</sup>

UREN, Université Virtuelle de Côte d'ivoire, Abidjan, CÔTE D'IVOIRE<sup>2</sup>

ImViA, Université de Bourgogne, Dijon, FRANCE<sup>2</sup>

LMI, Université Péléforo Gon Coulibaly, Korhogo, CÔTE D'IVOIRE<sup>3</sup>

**Abstract**—Cocoa cultivation is of immense importance to the people of Côte d'Ivoire. However, this culture is experiencing significant challenges due to diseases spread by various agents such as bacteria, viruses, and fungi, which cause considerable economic losses. Currently, the methods available to detect these cocoa diseases force farmers to seek the expertise of agronomists for visual inspections and diagnostics, a laborious and complex process. In the search for solutions, many studies have opted for using convolutional neural networks (CNNs) to identify diseases in cocoa pods. However, an essential advance is to develop hybrid approaches that combine the advantages of a CNN with sophisticated classification algorithms. This research stands out for its innovative contribution, combining MobileNetV2, a convolutional neural network architecture, with algorithms, such as Logistic Regression (LR), K Nearest Neighbors (KNN), Support Vector Machines (SVM), XGBoost, and Random Forest. The study was conducted in two distinct phases. First, each algorithm was evaluated individually, and then performance was measured when MobileNetV2 was merged with the algorithms mentioned. These hybrid approaches complement and amplify MobileNetV2's capabilities. To do so, they draw on MobileNetV2's inherent capabilities to extract key features and enhance information quality. By combining this expertise with the classification methods of these other models, hybrid approaches outperform individual techniques. Accuracy rates range from 72.4% to 86.04%. This performance amplitude underlines the effectiveness of the synergy between the extraction characteristics of MobileNetV2 and the classification skills of other algorithms.

**Keywords**—Cocoa pods diseases; MobileNetV2; classification algorithms; machine learning; hybrid method

## I. INTRODUCTION

Cocoa pod diseases are a significant problem for farmers and the cocoa industry [1]. These diseases can lead to yield loss, reduced cocoa bean quality, and higher production costs. Many methods are available to combat cocoa pod diseases, including pesticides, fungicides, and organic practices. However, these methods can be costly and difficult to apply.

Computer vision, generally based on machine learning and deep learning [2], is now being exploited for various agriculture, botany, and ecology tasks. These tasks include assessing the health of plants in our various crops. This new technology is proving necessary to improve crop yields in general. Researchers have already conducted numerous studies to identify disease risk factors, such as adverse weather

conditions [3], insect pests, and weeds. By identifying these risk factors, farmers can take steps to reduce or eliminate them, thus helping to prevent disease.

Indeed, Machine Learning and Deep Learning algorithms have shown enormous potential in agriculture by helping to develop effective treatments for crop diseases. They provide an innovative and powerful approach to combating crop diseases. Their ability to process complex data, detect anomalies early, and customize treatments significantly benefits crop health, agricultural sustainability, and food security [4] [5].

Applying Machine Learning (ML) and Deep Learning (DL) algorithms for detecting cocoa pod diseases offers tangible benefits for improving cocoa quality, increasing selling prices, and improving farmers' incomes. This approach supports farmer profitability and contributes to the sustainability of the cocoa sector by enhancing the quality and reputation of the cocoa produced.

More specifically, this study contributes to the sustainability of the cocoa sector by improving the quality and reputation of the cocoa produced. Early detection of disease enables farmers to take corrective action, reducing production losses and improving cocoa quality

Our research focuses mainly on identifying and detecting diseases affecting cocoa pods. It has several important implications. Firstly, it shows that the hybrid approach is promising for cocoa pod disease detection. Secondly, it suggests that CNNs can be used to improve the performance of conventional machine learning algorithms. Finally, it paves the way for new applications of machine learning and deep learning in the cocoa sector. Our contributions will be broken down into the following aspects:

As a first step, we will extract images from each pod using data augmentation techniques. This step will be crucial to establish the dataset that will serve as the basis for our study.

We will combine classic Machine Learning algorithms and a convolutional neural network (CNN) such as MobileNet.

We will explore hybrid approaches aimed at merging the capabilities of the MobileNetV2 network with the set of classic Machine Learning algorithms already in use.

Finally, we will evaluate the performances of different methods by analyzing the impact of integrating MobileNet in the fusion of the approaches.

The organization of our study is outlined as follows. Following the introductory section outlining the issue tackled in this paper (Section I), we delve into existing research on identifying specific cocoa diseases (Section II). Subsequently, we present and expound upon the methodology in Section III. Elaborate findings are disclosed in Section IV.

A comprehensive discussion is laid out in Section V, and ultimately, Section VI addresses the research goals and derives conclusions from the conducted study.

## II. RELATED WORK

Cocoa farming is a vital sector for the Ivory Coast and an essential resource for our farmers. Unfortunately, this crop is sometimes threatened by diseases that cause enormous losses for our farmers and Côte d'Ivoire, the world's leading cocoa producer. To preserve the gains made and further improve the production of this crop, a great deal of work has been carried out.

Godmalin et al. [6] carried out a study based on a deep learning algorithm to tackle the automatic classification of the state of a cocoa pod. Their experimental research method relies on a convolutional neural network for training. The model can classify three states of a given cocoa pod image: healthy, attacked by black pod disease, and shot by a pest. The results of their experiment showed an accuracy of 94%. However, the study did not assess the impact of weather conditions on the algorithm's performance. The algorithm's performance may vary according to weather conditions. Please do not revise any of the current designations. Sandra Kumi et al. have proposed a method using machine learning techniques to detect and diagnose two significant diseases affecting cocoa production, namely Swollen Shoot and Black Pod [7]. A mobile application with integrated ML techniques is offered to cocoa farmers to take a photo of the pod and upload it for diagnosis, which takes place on a cloud backend service. Four CNN models were built and trained for automatic disease detection and diagnosis. The results of this study showed that the MobileNet V2 SSD gave the best accuracy rate, with a score of 80%. However, the study was conducted in a controlled environment, and the algorithm may need to be more accurate in a real-life setting, where conditions can vary.

Amoako et al. [8] studied a model based on VGG19 to classify cocoa diseases using images. Then, other pre-trained models, such as VGG16 and ResNet50, are compared. Their study is a step in the right direction toward developing a technology that could positively impact the cocoa industry. With further research, it is possible to improve the accuracy of machine learning models and make them more applicable to a wide range of environments. This could lead to a significant reduction in losses due to cocoa diseases and an improvement in the quality of the cocoa produced. Basri et al. [9] proposed a study comparing the results of four feature extraction models in the case of early recognition of disease attacks on cocoa fruits. The image extraction models used include:

- Local binary pattern (LBP).
- Gray level co-occurrence matrix (GLCM).
- Hue saturation value (HSV).

Gray level co-occurrence histograms (GLCH).

In addition, the support vector machine (SVM) model was applied to the classification technique to measure the extraction results from the cocoa image dataset. SVM classification results revealed the best performance for HSV feature extraction for all types of SVM kernels applied (linear, RBF, and polynomial), with the highest accuracy being 80.95% for the RBF kernel.

Baba et al. [10] have studied a model based on image processing techniques for identifying early symptoms of pests and diseases in cocoa fruit from mobile applications. This research showed that the system's accuracy in recognizing cocoa fruit-shaped objects reached 100% in identifying cocoa fruit and 83% database for images of normal conditions at 83.75%, disease attack at 84.87%, and pest attack at 80.80%. Their study showed that image-processing techniques can accurately identify early symptoms of pests and diseases in cocoa fruit. However, the study also revealed areas for improvement in this approach, namely that many images can drag the model. In addition, the images need to be of high quality and well-lit. If this is not the case, the model may need to identify pests and diseases correctly. Gunawan et al. set up an expert system based on the Certainty Factor (CF) algorithm to identify the factor of cocoa pests and diseases that cannot be identified and prevented in advance [11]. Based on the results of the accuracy test, the proposed model can produce an accuracy rate of 86.67% in diagnosing pests and diseases on cocoa plants. However, the study used a simple algorithm, the certainty factor, which may not be able to detect all pests and diseases. Mohammad Yazdi et al. [12] have proposed a model for building a system to detect pest and disease types in cocoa pods. This study uses digital image processing techniques to extract color characteristics from digital images of cocoa pods. The method used to extract hue, saturation, and value (HSV) color characteristics and the classification algorithm used is K-Nearest Neighbor (KNN). One hundred fifty images were divided into 70% training data and 30% test data. Based on test results using k values of 5, 7, 11, and 13 in the restraint method, the best accuracy is 84.44% with a k = 5 value. However, the images used in the study are limited to a single type of cocoa tree, which could limit the generalizability of the results to other types of cocoa. It should be noted that the KNN algorithm is used to classify the images, but this simple algorithm may not detect the more complex pests and diseases. Godmalin et al. studied an experimental search method to train a convolutional neural network capable of classifying three levels of cocoa pod infection: low, moderate, and severe [13]. The results of this model achieved 91% accuracy in correctly classifying the condition of cocoa pods.

Overall, this study is a step in the right direction. Still, more work needs to be done to develop automatic detection methods for cocoa pod infection that are reliable and usable in real-life conditions.

## III. MATERIAL AND METHOD

### A. Materials

The database used for our study is Cocoa Diseases [14], an online database launched in 2020 by the Autonomous

University of Bucaramanga, Colombia. It comprises 312 images with 1,591 tagged objects belonging to three classes: healthy, phytophthora, and monilia. The experiments used Python programming on a THINPAD laptop with an Intel(R) Core i7-10700 processor running at 2.90 GHz, 32 GB memory, and a 512 GB SSD hard disk.

## B. Methods

The machine learning methods we used in our study are as follows.

1) *SVM* : A support vector machine (SVM) is a supervised machine learning algorithm [15] that can be used to solve classification and regression problems. SVMs are a generalization of linear classifiers. The principle behind SVMs is to find a hyperplane that optimally isolates data from two classes. The optimal hyperplane corresponds to the one that optimizes the margin between the data of the two classes [16]. The data closest to the hyperplane are called support vectors. SVMs are effective for solving classification and regression problems with non-linear data. They also adapt well to new data. Effective for solving classification and regression problems with non-linear data, it is an algorithm capable of generalizing well to new data. SVMs are used in various applications, such as image classification, facial recognition, object detection, text classification, and regression.

2) *Random forest* : The Random Forest algorithm is a supervised machine learning algorithm that uses a set of decision trees to make predictions [17]. It is known for its accuracy and robustness and is used in various fields, including classification, regression, and anomaly detection. The algorithm starts by creating a set of decision trees. The number of trees in the set is a parameter the user can adjust. Each decision tree is trained on a random subset of the training data. Data that is not used to train a decision tree is called test data. The algorithm then uses the test data to evaluate the accuracy of each decision tree. The algorithm then averages the predictions of all the decision trees to make a final prediction. The advantage of the Random Forest algorithm is that it can make more accurate predictions than the individual decision trees. The decision trees are uncorrelated, so they don't make the same errors. The Random Forest algorithm is also robust to noisy data. This is because the decision trees in the ensemble can adapt to variations in the data. The Random Forest algorithm is a powerful tool that can be used to solve various machine learning problems. It is renowned for its accuracy, robustness, and ease of use.

3) *XGBoost*: XGBoost is an improved model of the Gradient Boost algorithm. This machine-learning algorithm can solve common commercial problems using minimal resources [18]. Extreme Gradient is a method used to reduce the number of errors in predictive data analysis. XGBoost is an assembly of decision trees (weak learners) that predict residuals and correct mistakes in previous decision trees. The unique feature of this algorithm lies in the decision tree used. This recently introduced machine learning algorithm has proved very powerful for modeling complex processes in other

research fields [19]. It is a robust algorithm that can solve various machine learning problems. It is known for its accuracy, speed, and flexibility.

4) *KNN*: The KNN (k-nearest neighbor) algorithm is a supervised learning algorithm used for classification and regression [5]. It works by calculating the distance between an unknown point and known points in the training dataset. The k points closest to the unknown point are then used to predict the class or value of the unknown point. The KNN algorithm is non-parametric, meaning that it makes no assumptions about the distribution of the data [2]. It is also a computationally inexpensive algorithm, making it suitable for large quantities of data. It works by calculating the distance between an unknown point and known points in the training data set. The k points closest to the unknown point are then used to predict the class or value of the unknown point. The result is a computationally inexpensive algorithm, making it suitable for large amounts of data.

5) *Logistic regression* is a supervised learning algorithm [20] used to predict the probability of an observation belonging to a particular class. It is often used for binary and multiclass classification [21]. It is a linear model, which means that the probability of class membership is modeled as a linear combination of the input variables. The probability function is a logistic function, which is a non-linear function that takes a value between 0 and 1 [22]. Logistic regression is a robust algorithm that can solve many problems. The algorithm begins by calculating the model weights. Weights are coefficients that measure the importance of each input variable. The weights are then used to calculate the probability of belonging to each class. The observation is classified in the category with the highest probability. The choice of weights is important for model performance. Weights can be optimized using a technique called gradient descent. Gradient descent is an optimization technique for finding weights that minimize model error. Logistic regression is a robust algorithm that can solve many problems. It is easy to implement and understand, making it popular with data scientists.

6) *MobileNetV2*: MobileNet V2 is a lightweight convolutional neural network model developed by Google AI [23]. It was first introduced in a research paper published in 2018. MobileNet V2 is based on the original MobileNet architecture but introduces several improvements, namely the use of a new convolution block called "bottleneck" and that of a learning technique called "transfer learning". MobileNet V2 has been designed for mobile devices such as smartphones and tablets. It can achieve performance comparable to larger models while being much lighter and more energy-efficient [24]. The bottleneck convolution block is a convolution block [25] that reduces the width and height of the input signal while retaining the depth. This reduces the number of model parameters while maintaining performance. MobileNet V2 is a powerful and versatile model that can be used for various tasks, such as image classification, object detection, and facial



recognition. It is lightweight and energy-efficient, making it ideal for mobile devices.

The methodology adopted in this study is based on several well-defined phases, providing a solid and rigorous foundation for our research. Detection techniques are studied to enable the system to detect and identify cocoa pod diseases. A comprehensive outline of these stages is provided in Fig. 1:

Our methodological approach is based on a series of carefully developed steps:

- Database preparation: In this phase, each image was segmented to identify each pod in different images precisely. To improve the quality of our data, each series of images associated with a pod was subjected to a data augmentation step.
- Data Augmentation: Data augmentation is an essential element of our approach. In this step, random transformations were applied to images. Each image was subjected to random horizontal and vertical flipping, allowing left and right as well as top and bottom inversions. In addition, a random rotation was applied to each image, with a maximum rotation angle of 0.4 radians (approx. 23 degrees) clockwise.
- Data Division: For the training and evaluation of our models, we have divided our data into three sets: the

training set, the validation set, and the test set. This division enables us to evaluate and validate the performance of our models.

- Using Various Algorithms: To tackle our pod disease detection task, we adopted a varied approach using traditional Machine Learning algorithms such as SVM, Random Forest, KNN, Logistic Regression and XGBoost, and deep methods based on the MobileNetV2 network.
- Hybrid Methods: In recognition of the complementary power of the approaches, we have also explored hybrid methods that merge the potential of the MobileNetV2 network with the set of Machine Learning algorithms mentioned above.
- Performance Evaluation: To assess the effectiveness of each model, we carefully evaluated their ability to detect pod diseases. This crucial phase enabled us to quantify and compare the performance of each approach.

Our methodology is built on a solid foundation, from careful data preparation to thorough model evaluation to detect pod diseases accurately.

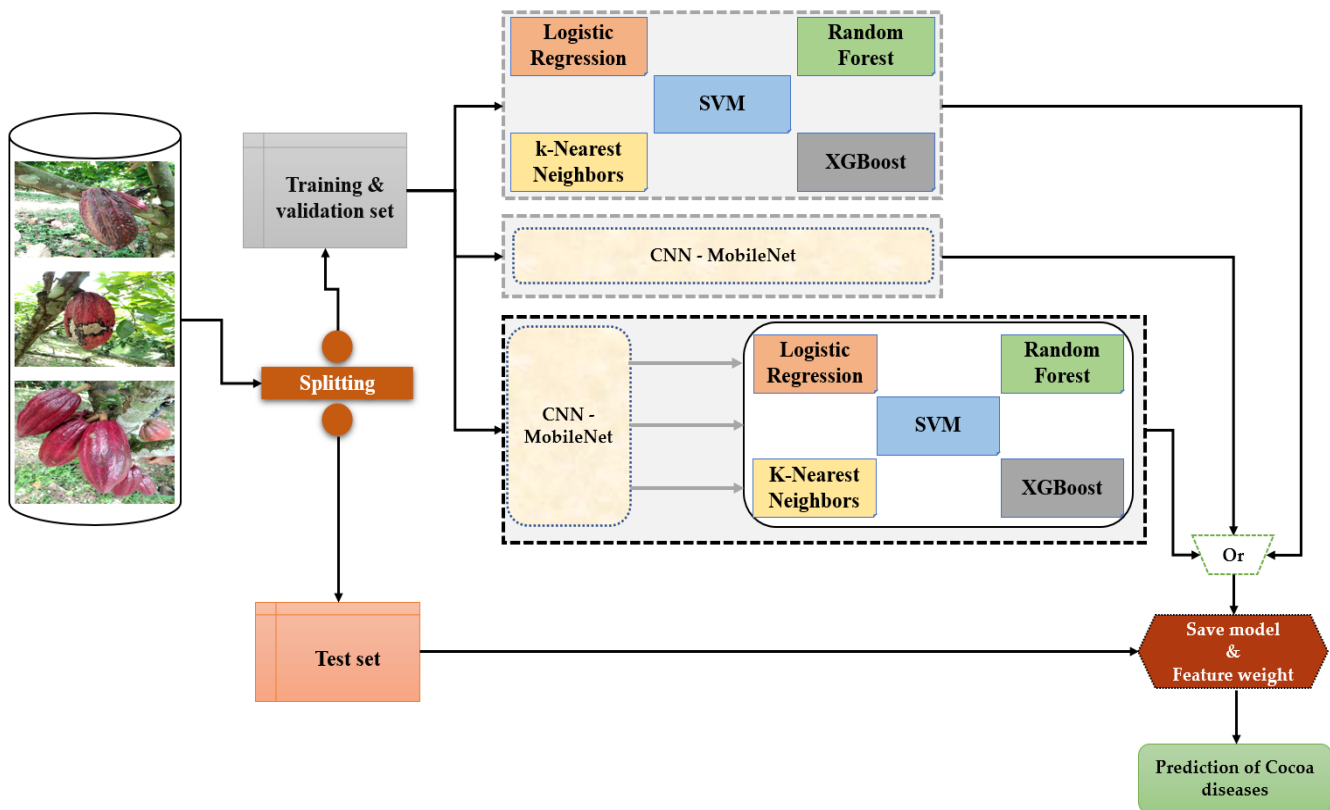


Fig. 1. Illustration of our methodology.

C. Evaluation Metrics

In evaluating the outcomes of our investigation, we employed various metrics. We included accuracy, precision, recall, F1 score, and Matthew's correlation coefficient (MCC) within this set of measures. The differential equations are outlined as follows:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1\ score = 2 * \frac{Precision*Recall}{Precision+Recall} \quad (4)$$

$$MCC = \frac{TP*TN-FP*FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (5)$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (6)$$

The Receiver Operating Characteristic (ROC) curve is a graphical representation showcasing the performance of a binary classification model. It illustrates the relationship between the true positive rate (sensitivity) and the false positive rate across various classification thresholds. An ideal ROC curve aligns with the upper left corner of the graph, indicating high sensitivity and specificity.

The confusion matrix is a tabular summary that encapsulates the outcomes of a classification model's predictions. It assesses the model's predictions against the actual values in the dataset, categorizing them into four groups: true positives, true negatives, false positives, and false negatives. The confusion matrix compares the model's precision, recall, specificity, and accuracy.

IV. RESULTS

Our findings will be presented in two distinct sections, each exploring a specific area in depth:

A. Individual Algorithm Performance

The first part of our results will be dedicated to evaluating and comparing the performance of the various Machine Learning algorithms we have used. These algorithms include SVM, Random Forest, KNN, Logistic Regression, XGBoost, and the MobileNetV2 Deep Learning algorithm. Each algorithm has been tested and analyzed, enabling us to determine its specific effectiveness in detecting pod diseases. We will evaluate each algorithm's key metrics, giving us a comprehensive view of its ability to meet this challenge. Table I shows the metric measurements for each model.

Fig. 2 shows the confusion matrices and shows the ROC curves for the three best accuracies.

TABLE I. CASE OF INDIVIDUAL ALGORITHM METRIC RESULT

Models	Accuracy (%)	Precision (%)	F1 score (%)	Recall (%)	MCC (%)	MSE
SVM	67,09	66,78	66,78	67,09	48,49	0,81
Random Forest	59,53	63,28	63,28	64,17	43,54	0,91
KNN	46,96	39,54	39,54	46,96	18,07	1,47
Logistic Regression	54,65	53,71	53,71	54,65	29,06	1,08
XGBoost	68,61	67,79	67,79	68,61	50,95	0,77
MobileNetV2	<b>81,66</b>	<b>81,4</b>	<b>81,4</b>	<b>81,65</b>	<b>72,09</b>	<b>0,4</b>

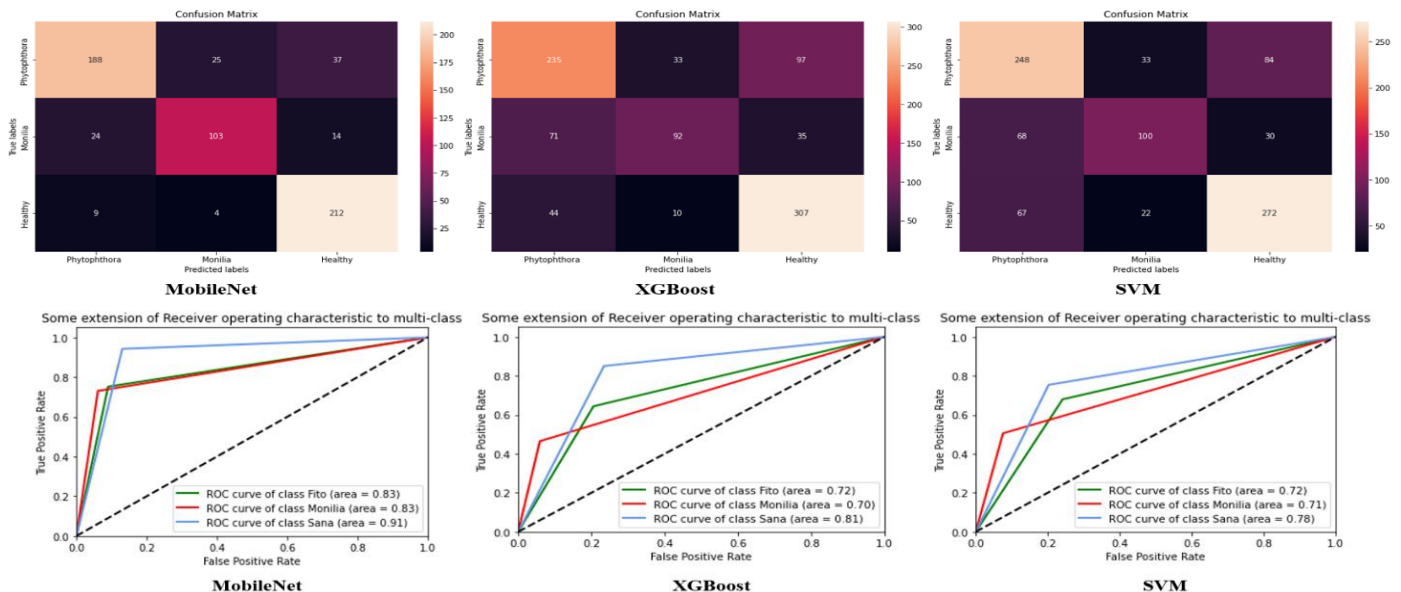


Fig. 2. Confusion matrix and ROC curve of individual algorithm.

The analysis reveals significant variations in the performance of each model. The MobileNetV2 model displays the highest accuracy at 81.66%, closely followed by logistic regression and SVM. However, when considering measures such as recall and F1 score, MobileNetV2 stands out, suggesting its ability to identify true positives well. Furthermore, the MCC, which assesses the overall prediction quality, puts MobileNetV2 in the lead with 72.09%.

On the other hand, KNN's performance could be better regarding precision, recall, and F1 score, suggesting difficulty discerning true positives. The lowest MCC also corroborates this among the models tested.

In sum, these results highlight the superiority of MobileNetV2 in terms of overall performance. Still, they also underline each model's specific strengths and weaknesses in the pod disease detection task.

### B. Hybrid Method Performance

The other part of our results will focus on an in-depth examination of hybrid methods. This section will explore the synergies between the Deep Learning-based MobileNetV2 network and the previously mentioned machine learning algorithms. The performance of these hybrid methods will be rigorously analyzed, revealing whether their combination can achieve even higher detection levels. This evaluation will give us a significant perspective on the added value provided by the fusion of these approaches.

Table II shows the metric measurements for each model.

Fig. 3 shows the confusion matrices and shows the ROC curves for the three best accuracies

TABLE II. CASE OF HYBRID METHOD ALGORITHM METRIC RESULT

Models	Accuracy (%)	Precision (%)	F1 score (%)	Recall (%)	MCC (%)	MSE
<b>MobileNetV2 - SVM</b>	<b>86,04</b>	<b>85,88</b>	<b>85,88</b>	<b>86,03</b>	<b>78,53</b>	<b>0,33</b>
MobileNetV2 - RF	77,60	76,44	76,44	77,59	65,6	0,49
MobileNetV2 - KNN	72,4	71,59	71,59	72,4	61,03	0,76
MobileNetV2 - LR	85,88	85,71	85,71	85,87	78,29	0,31
MobileNetV2 - XGBoost	79,38	78,99	78,99	79,38	68,11	0,45
MobileNetV2	81,66	81,4	81,4	81,65	72,09	0,4

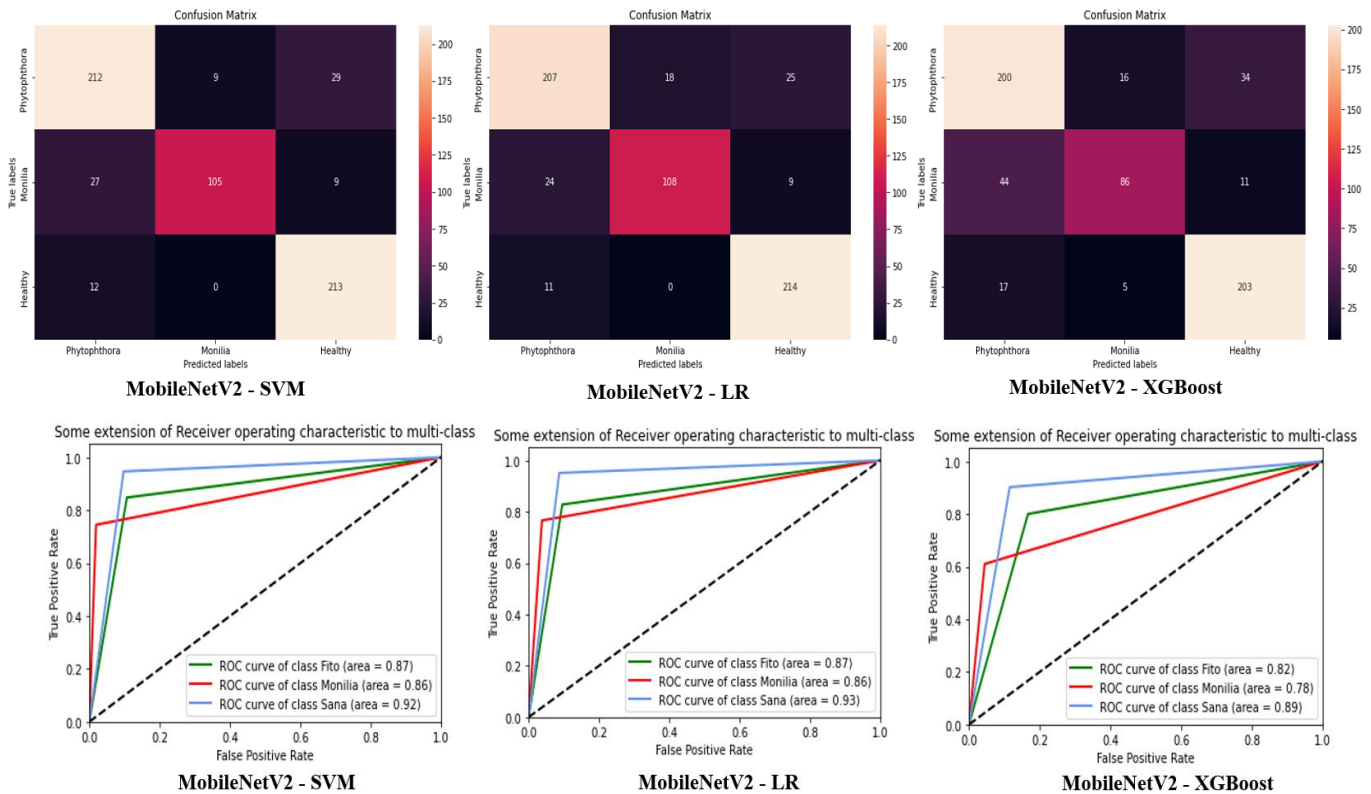


Fig. 3. Confusion matrix and roc curve of hybrid method algorithm.

Analysis of the table above highlights the performance and trends of the various models, including combinations of MobileNetV2 with multiple algorithms and the performance of MobileNetV2 as a stand-alone model. Each combination presents distinct characteristics regarding the accuracy, precision, F1 score, recall, Matthews Correlation Coefficient (MCC), and Mean Square Error (MSE).

The combination of MobileNetV2 with SVM is the most efficient, with an accuracy of 86.04%. This means that it can correctly classify 86.04% of images. Precision, F1 score, and recall are also high, indicating that the model can accurately identify true positives and minimize false positives and false negatives. The combination of MobileNetV2 with Random Forest is also effective, with an accuracy of 77.60%. However, the precision, F1 score, and recall measures are slightly lower compared with the MobileNetV2-SVM combination. This means the model is less accurate at identifying true positives but more likely to minimize false positives.

The combination of MobileNetV2 with KNN is less effective, with an accuracy of 72.4%. Precision, F1 score, and recall measures are all relatively similar, indicating that the model can detect true positives and negatives with comparable precision.

The combination of MobileNetV2 with LR is effective, with an accuracy of 85.88%. The high accuracy associated with this combination indicates an ability to avoid false positives. The high F1 and recall scores show that the model can effectively detect true positives. The combination of MobileNetV2 with XGBoost is comparable to other hybrids, with an accuracy of 79.38%. Precision, F1 score, and recall demonstrate a balanced ability to identify true positives and minimize false positives and negatives. As a stand-alone model, MobileNetV2 is effective, with an accuracy of 81.66%. Other measures, including precision, F1 score, recall, and MCC, also show balanced performance.

## V. DISCUSSION

The results reveal key aspects concerning the performance of the different models evaluated for pod disease detection. These results highlight distinct strengths and weaknesses of each model, providing important information to guide the optimal mode selection.

In terms of individual performance, MobileNetV2 had the best overall performance of all the models evaluated. Its high accuracy of 81.66% indicates its ability to deliver accurate predictions. In addition, its high F1 score and balanced recall (81.4% and 81.65%, respectively) indicate its ability to identify

true positives while maintaining a balance with false negatives. The high MCC of 72.09% reflects an excellent correlation between predictions and actual observations. Although MobileNetV2 performs well, it may require higher computing resources due to its complex architecture.

**SVM and Logistic Regression:** These traditional Machine Learning models show stable and balanced performance, with precision, F1 score, and recall rates around 67%. They can be considered reliable options for the detection of pod diseases.

However, their inability to achieve the same levels of accuracy as MobileNetV2 may suggest limitations in their ability to capture subtle image features.

**Random Forest and XGBoost:** These ensemblistic models show comparable results, with F1 scores and balanced recalls. They show a certain robustness in detecting true positives.

However, their slightly lower accuracy could mean that they tend to identify some false positives.

Although KNN has the lowest accuracy among the models evaluated, its relatively high MCC reveals a specific relevance in detecting pod diseases.

However, its low recall and F1 score underline its difficulty correctly identifying all true positives.

Model performance varies according to method and architecture. While MobileNetV2 stands out for its high accuracy and correlation, traditional Machine Learning models offer a solid and stable option. By considering the strengths and weaknesses of each model, informed decisions can be made to maximize the quality of pod disease detection.

Fig. 4 presents the histogram of the metrics of classical algorithms.

Combining MobileNetV2 with models such as SVM, LR, or XGBoost improves overall performance compared with MobileNetV2 alone. This is due to how these models complement MobileNetV2's capabilities, leveraging the power of MobileNetV2's computer vision and the more traditional classification methods of these other models.

The combination with SVM seems particularly effective in terms of Accuracy and MCC, showing how the SVM approach can enhance MobileNetV2's classification capabilities. MobileNetV2 combined with KNN or Random Forest also shows improvements, but performance is still inferior to that obtained with SVM or XGBoost.

Fig. 5 presents the metric histogram of the hybrid methods.

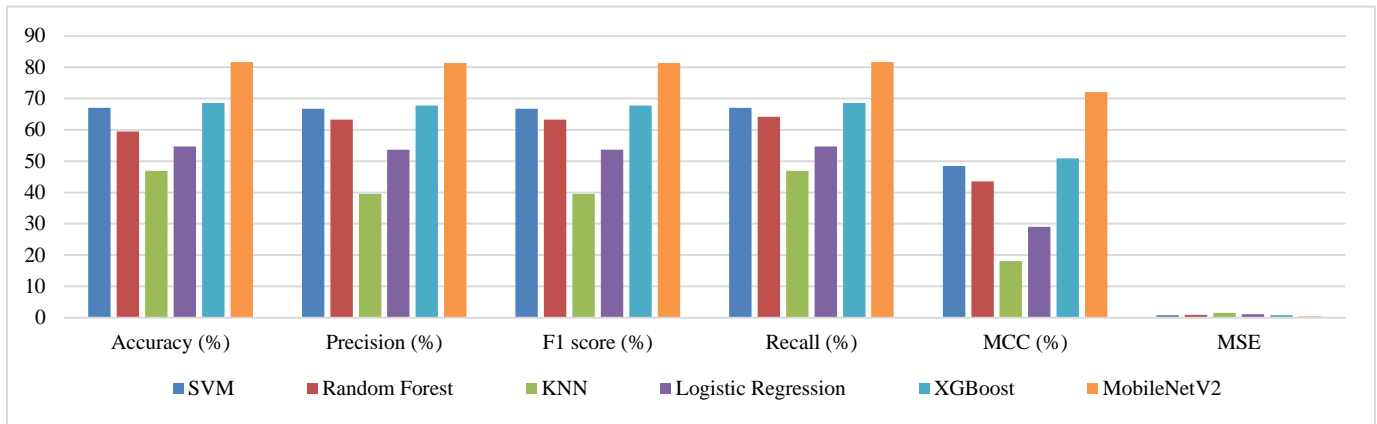


Fig. 4. Model performance histogram of the case of individual algorithm.

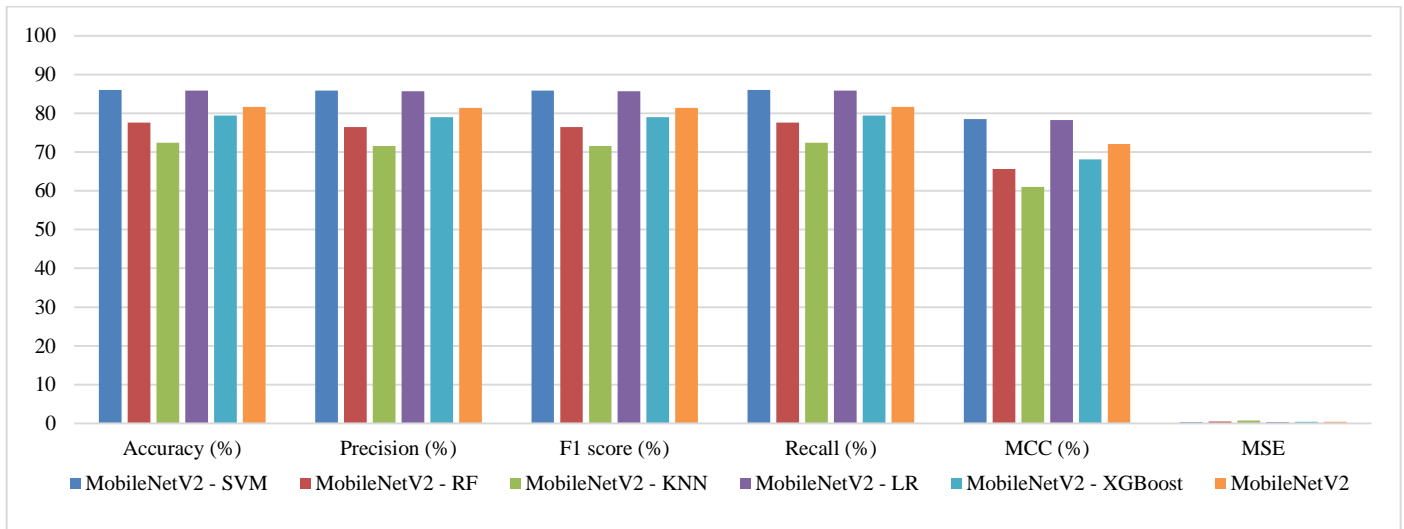


Fig. 5. Model performance histogram of the case of hybrid method algorithm.

## VI. COMPARISON WITH EXISTING APPROACHES

The results of our work exceeded those of several related studies, including previous research on the detection and identification of cocoa pests and diseases.

Table III shows the results. The results clearly show that our model outperforms that of the other two authors.

TABLE III. COMPARISON OF RESULTS WITH PREVIOUS RESEARCH

Method	Accuracy (%)
Basri B, et al [26]	82.50
RY Montesino et al[27]	83
Our method	86.04

## VII. CONCLUSION

This study represents a significant step forward in the fight against cocoa diseases in Côte d'Ivoire. By combining the intelligence of MobileNetV2's convolutional neural networks with the advanced classification capabilities of algorithms such as Logistic Regression, K-nearest Neighbors, Support Vector

Machines, XGBoost, and Random Forest, we have succeeded in significantly improving the accuracy of disease detection.

The results obtained are promising, highlighting the effectiveness of hybrid approaches in tackling complex agricultural problems. This innovative method offers an alternative to traditional methods of laborious visual inspection, enabling farmers to take more targeted measures to protect their crops.

However, it is important to note that further research is needed to refine these hybrid approaches and adapt them to the seasonal and environmental variations to which cocoa is exposed. In addition, additional validation on varied datasets and real field conditions would reinforce the conclusions' robustness. This study lays the foundations for a new era in cocoa disease monitoring and management, with potentially positive implications for farmers, local economies, and food security. As technology continues to evolve, this approach could serve as a model for solving similar problems in other areas of agriculture and beyond.

The prospects of this study could pave the way for the development of new applications of machine learning and deep

learning in the cocoa sector. Indeed, the hybrid approach developed by the researchers proved effective in detecting disease in cocoa pods. This approach could be used to develop new applications, such as a mobile diagnostic tool that would enable farmers to detect cocoa pod diseases on the spot or to monitor the spread of cocoa diseases.

#### REFERENCES

- [1] « Renforcer La Compétitivité de La Production de Cacao Et Augmenter Le Revenu Des Producteurs de Cacao en Afrique de L'ouest Et en Afrique Centrale | PDF | Solides de cacao | Afrique », Scribd. <https://fr.scribd.com/document/505690969/2017-07-Renforcer-la-competitivite-de-la-production-de-cacao-et-augmenter-le-revenu-des-producteurs-de-cacao-en-Afrique-de-l-Ouest-et-en-Afrique-centr> (consulté le 24 août 2023).
- [2] « Boughaba\_Boukhris.pdf ». Consulté le: 24 août 2023. [En ligne]. Disponible sur: [https://dspace.univ-ouargla.dz/jspui/bitstream/123456789/17195/1/Boughaba\\_Boukhris.pdf](https://dspace.univ-ouargla.dz/jspui/bitstream/123456789/17195/1/Boughaba_Boukhris.pdf)
- [3] « Réseaux Des Capteurs Sans Fil Intelligent Pour La Détection Des Mauvaises Herbes Dans Les Applications Agricoles ». <https://theses-algerie.com/2729392144012869/memoire-de-master/universite-larbi-tebessi---tebessa/r%C3%A9seaux-des-capteurs-sans-fil-intelligent-pour-la-d%C3%A9tection-des-mauvaises-herbes-dans-les-applications-agricoles> (consulté le 24 août 2023).
- [4] T. Domingues, T. Brandão, et J. C. Ferreira, « Machine Learning for Detection and Prediction of Crop Diseases and Pests: A Comprehensive Survey », *Agriculture*, vol. 12, no 9, Art. no 9, sept. 2022, doi: 10.3390/agriculture12091350.
- [5] K. J. Ayikpa, D. Mamadou, P. Gouton, et K. J. Adou, « Experimental Evaluation of Coffee Leaf Disease Classification and Recognition Based on Machine Learning and Deep Learning Algorithms », *Journal of Computer Science*, vol. 18, no 12, p. 1201-1212, déc. 2022, doi: 10.3844/jcssp.2022.1201.1212.
- [6] R. A. Godmalin, C. J. Aliac, et L. Feliscuzo, « Classification of Cacao Pod if Healthy or Attack by Pest or Black Pod Disease Using Deep Learning Algorithm », in 2022 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAET), Kota Kinabalu, Malaysia: IEEE, sept. 2022, p. 1-5. doi: 10.1109/IICAET55139.2022.9936817.
- [7] S. Kumi, D. Kelly, J. Woodstuff, R. K. Lomotey, R. Orji, et R. Deters, « Cocoa Companion: Deep Learning-Based Smartphone Application for Cocoa Disease Detection », *Procedia Computer Science*, vol. 203, p. 87-94, 2022, doi: 10.1016/j.procs.2022.07.013.
- [8] P. Y. O. Amoako, G. Cao, et J. K. Arthur, « An Image-Based Cocoa Diseases Classification Based on an Improved Vgg19 Model », Springer Books, p. 711-722, 2023.
- [9] Basri, Indrabayu, A. Achmad, et I. S. Areni, « Comparison of Image Extraction Model for Cocoa Disease Fruits Attack in Support Vector Machine Classification », in 2022 International Conference on Electrical and Information Technology (IEIT), Malang, Indonesia: IEEE, sept. 2022, p. 46-51. doi: 10.1109/IEIT56384.2022.9967910.
- [10] B. Baba, R. Tamin, Indrabayu, I. Areni, et H. A. Karim, « MOBILE IMAGE PROCESSING APPLICATION FOR CACAO'S FRUITS PEST AND DISEASE ATTACK USING DEEP LEARNING ALGORITHM », *ICIC Express Letters*, vol. 14, p. 1025-1032, oct. 2020, doi: 10.24507/icicel.14.10.1025.
- [11] R. D. Gunawan, I. Ahmad, Parjito, H. Anggono, E. H. Vernando, et F. Jaya, « Optimization of Certainty Factor Algorithm to Overcome Uncertainty in Expert System Identification of Pests and Diseases of Cocoa », in 2022 2nd International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS), Yogyakarta, Indonesia: IEEE, nov. 2022, p. 310-315. doi: 10.1109/ICE3IS56585.2022.10010123.
- [12] Mohammad Yazdi Pusedan, Syahrullah, Merry, et Ahmad Imam Abdullah, « k-Nearest Neighbor and Feature Extraction on Detection of Pest and Diseases of Cocoa », *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 6, no 3, p. 471-480, juill. 2022, doi: 10.29207/resti.v6i3.4064.
- [13] R. A. Godmalin, C. J. Aliac, et L. Feliscuzo, « Cacao Pod Infection Level Classification Using Transfer Learning », in 2023 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream), Vilnius, Lithuania: IEEE, avr. 2023, p. 1-6. doi: 10.1109/eStream59056.2023.10135062.
- [14] « Cocoa Diseases (YOLOv4) ». <https://www.kaggle.com/datasets/serranosebas/enfermedades-cacao-yolov4> (consulté le 24 août 2023).
- [15] A. K. Jean, M. Diarra, B. A. Bakary, G. Pierre, et A. K. Jérôme, « Application based on Hybrid CNN-SVM and PCA-SVM Approaches for Classification of Cocoa Beans », *IJACSA*, vol. 13, no 9, 2022, doi: 10.14569/IJACSA.2022.0130927.
- [16] C. Jabour, « Estimation de la résistance coronarienne par analyse du réseau vasculaire du fond de l'œil », phdthesis, Institut National des Sciences Appliquées de Lyon, 2023. Consulté le: 24 août 2023. [En ligne]. Disponible sur: <https://hal.science/tel-04104565>
- [17] D. Jacob, R. Tievant, L. Cervoni, et M. Roudesli, « Prédiction des blessures au Foot 5 à l'aide d'une méthode de machine learning », *Journal de Traumatologie du Sport*, juin 2023, doi: 10.1016/j.jts.2023.06.001.
- [18] B. Pan, « Application of XGBoost algorithm in hourly PM2.5 concentration prediction », vol. 113, p. 012127, févr. 2018, doi: 10.1088/1755-1315/113/1/012127.
- [19] B. M. Kouassi, V. Monsan, A. B. Ballo, K. J. Ayikpa, D. Mamadou, et K. J. Adou, « Application of the Learning Set for the Detection of Jamming Attacks in 5G Mobile Networks », *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 14, no 6, Art. no 6, 30 2023, doi: 10.14569/IJACSA.2023.0140676.
- [20] R. Benzitouni, L. Merabet, et S. Zertal, « La Proposition d'une Architecture basée Deep Learning pour la prédiction des maladies cardiaques dans un environnement IoT », 2022, Consulté le: 24 août 2023. [En ligne]. Disponible sur: <http://localhost:8080/xmlui/handle/123456789/14486>
- [21] C. Moussaoui et A. (Encadreur) Mellit, « Conception et réalisation d'un système semi-automatique de diagnostic des générateurs photovoltaïques à l'aide de l'apprentissage profond et de l'imagerie thermique. », Thesis, Université de Jijel, 2022. Consulté le: 24 août 2023. [En ligne]. Disponible sur: <http://dspace.univ-jijel.dz:8080/xmlui/handle/123456789/12453>
- [22] « (1) (PDF) Evaluation de l'impact de la territorialisation du Plan Maroc Vert sur le niveau de vie des phoeniculteurs des zones oasiennes de la région Drâa- Tafilalet à travers une régression logistique binaire ». [https://www.researchgate.net/publication/371812137\\_Evaluation\\_de\\_l'im pact\\_de\\_la\\_territorialisation\\_du\\_Plan\\_Maroc\\_Vert\\_sur\\_le\\_niveau\\_de\\_vi \\_des\\_phoeniculteurs\\_des\\_zones\\_oasiennes\\_de\\_la\\_region\\_Draa-\\_Tafilalet\\_a\\_travers\\_une\\_regression\\_logistique\\_binaire](https://www.researchgate.net/publication/371812137_Evaluation_de_l'im pact_de_la_territorialisation_du_Plan_Maroc_Vert_sur_le_niveau_de_vi _des_phoeniculteurs_des_zones_oasiennes_de_la_region_Draa-_Tafilalet_a_travers_une_regression_logistique_binaire) (consulté le 24 août 2023).
- [23] P. N. Srinivasu, J. G. Sivasai, M. F. Ijaz, A. K. Bhoi, W. Kim, et J. J. Kang, « Classification of skin disease using deep learning neural networks with mobilenet v2 and lstm », *Sensors*, vol. 21, no 8, avr. 2021, doi: 10.3390/s21082852.
- [24] A. Ballo, M. Diarra, A. Kacoutchy Jean, K. Yao, A. Assi, et K. Fernand, « Automatic Identification of Ivorian Plants from Herbarium Specimens using Deep Learning », *International Journal of Emerging Technology and Advanced Engineering*, vol. 12, p. 56-66, mai 2022, doi: 10.46338/ijetae0522\_07.
- [25] « Sensors | Free Full-Text | Accelerating 3D Convolutional Neural Network with Channel Bottleneck Module for EEG-Based Emotion Recognition ». <https://www.mdpi.com/1424-8220/22/18/6813> (consulté le 24 août 2023).
- [26] Basri, B., Indrabayu, I., Achmad, A., & Areni, I. S. (2023). A Review of Image Processing Techniques for Pest and Disease Early Identification Systems on Modern Cocoa Plantation. *International Journal of Computing and Digital Systems*, 13(1), 1-1.
- [27] RY Montesino, JA Rosales-Huamani et JL Castillo-Sequera, « Détection de phytophthora palmivora dans les fruits du cacao avec apprentissage profond », 2021 16e Conférence ibérique sur les systèmes et technologies d'information (CISTI) , Chaves, Portugal, 2021, pp. 1-4 , est ce que je : 10.23919/CISTI52073.2021.9476279.

# Wireless Capsule Endoscopy Video Summarization using Transfer Learning and Random Forests

Parminder Kaur<sup>1</sup>, Dr. Rakesh Kumar<sup>2</sup>

Dept. of Computer Science, Dr. B.R. Ambedkar Government College, Kaithal, Haryana, India<sup>1</sup>

Dept. of Computer Science and Applications, Kurukshetra University, Kurukshetra, Haryana, India<sup>2</sup>

**Abstract**—Wireless Capsule Endoscopy (WCE) is a diagnostic technique for identifying gastrointestinal diseases and abnormalities. Gastroenterologists face a considerable challenge when reviewing a lengthy video to identify a disease. The solution to this problem is generating an automated video summarization technique that generates the WCE Video summaries. This paper presents a Video Summarization technique that summarizes the WCE video. The proposed method uses transfer learning and a Random Forest classifier. Using a computationally light and pre-trained MobileNetV2 for feature extraction helped deliver results quickly. Managing small datasets and mitigating the overfitting risk was effectively addressed using Random Forest. The Random Forest's hyperparameters are optimized through the use of Bayesian optimization. The approach proposed has achieved an accuracy of 98.75% in disease prediction while significantly reducing the viewing time for the video summary. Furthermore, it has attained an average F-Score of 0.98, demonstrating its efficacy and reliability.

**Keywords**—Bayesian optimization; capsule endoscopy; MobileNetV2; random forest classifier; transfer learning

## I. INTRODUCTION

Wireless Capsule Endoscopy [1-3] is a technology used for performing the endoscopy of a patient to diagnose an illness. A pill-sized capsule camera captures the video of the gastrointestinal tract and assists the doctor in diagnosing any gastrointestinal abnormality. Capsule Endoscopy is effortless and does not interfere with a person's routine; due to this, it is getting more popular. However, specific challenges are associated with it - the battery may get low, and the capsule may get stuck in the gastrointestinal tract, and analyzing the long endoscopy video to identify the disease. An automated process for analyzing and generating the WCE video summary may save the doctor's time spent analyzing the video. Many researchers have utilized machine learning and deep learning methods to summarize WCE videos. Numerous studies aim to identify a particular ailment from a WCE video, like abnormal bleeding, tumor, polyp, or ulcer. However, if a problem-specific method is adopted, that proposed framework can only be applied to identifying a specific type of disease. One such approach was adopted in [4] for polyp detection having a lower false-positive rate. Since polyps are rounded or curved growths in the colon, the author also utilized the shape and texture features for polyp identification and localization. Similarly, [5] developed a classification method for polyp detection by using the textural characteristics of polyps and an improved bag of features. [6] Used Uniform Local Binary Pattern (LBP) to detect the polyps' texture. A method to detect Crohn's disease

using a deep convolutional network is proposed in [7]. According to a study [8], two critical factors can aid in detecting tumors: color and textural features. To achieve this, the SVM utilizes the LBP operator's feature maps. This approach has proven to be quite effective in accurately identifying tumors. It achieved an accuracy of 92.4% in detecting tumors. A saliency map-based ulcer detection method was proposed in [9]. A multilevel approach was used for detecting saliency and identifying ulcers. In [10], a CNN-based approach for bleeding detection is proposed. The CNN developed has a low complexity because the input to the CNN is a single patch, and it outputs a segmented patch of the same size. An approach for detecting multiple bleeding detection was proposed in [11]. For detecting the small intestine lesions, [12] used AlexNet. The model achieved an accuracy of over 95.16%. [13] Proposed a technique for lesion detection using the high-level features extracted by ResNet50 [14] and InceptionV4. The lesion and non-lesion frames are classified by using the SVM classifier.

Specific video summarization approaches focus on keyframe extraction. A key frame is the most informative and relevant frame. [15], used a keyframe extraction approach for video summarization. The irrelevant frames are discarded in the first phase of image quality assessment. From the remaining frames, keyframes are extracted using low-level and deep features. Another keyframe extraction-based video summarization technique is proposed in [16]. Convolutional autoencoder is used to extract high-level features and shot boundary detection. A shot is part of a video having similar content. Motion profiles are used to extract keyframes from each shot. In [17], temporal segmentation of the video into shots was accomplished with the Prune Exact Linear Time (PELT) algorithm, and the high-level features were extracted by using pre-trained VGG19. A temporal segmentation of endoscopy video for detecting abnormal video segments was proposed by [18]. The abnormal shot covers any gastrointestinal abnormality. To identify the abnormality a Graph Convolutional Network (GCNN) was used. A keyframe selection strategy for polyp identification by utilizing the depth information of polyps is proposed in [19]. One approach for constructing keyframes of endoscopic videos uses pre-trained InceptionV3 to create feature maps of WCE images, which are then fed into a K-means algorithm implemented in [20].

A machine learning model's efficiency largely depends upon the amount of data used to train the model. If the training dataset is prominent, the model may learn adequately; otherwise, it may overfit, leading to inaccurate results with the

test data. Nevertheless, in certain situations and domains, immense amounts of training data are unavailable; training a deep learning model for that problem becomes tedious. Transfer learning is the solution in such cases. Transfer learning is transferring knowledge from a learned model to a new one. However, both models should perform similar kinds of tasks. Applying transfer learning has several advantages over making a model learn from scratch. Training a model from scratch is time-consuming and requires tremendous training data. Moreover, there is no point in training a model from scratch if it performs a task similar to that performed by some other model.

The main objective of this research paper is to create a computationally efficient model that delivers precise results quickly. The proposed solution aims to overcome the following challenges:

1) One of the challenges in developing a machine-learning model for WCE is the need for adequate training data. With sufficient data, it is possible to train the model effectively. However, transfer learning is a solution to this problem. One can overcome the data shortage by leveraging pre-trained models with their weights and still successfully train a model.

2) Timely delivery of final results is crucial for the diagnostic procedure of endoscopy. Any delays are deemed unacceptable and can have serious consequences. MobileNetV2 [21] is a lightweight model in terms of computational requirements. Despite this, it can provide accurate solutions on time.

3) Machine learning models need a lot of computational resources. However, a model that requires significantly less computational resources and is so computationally light that it can be used in a mobile device is a favored solution.

This paper proposes a WCE video summarization technique that extracts the frames' deep features using the MobileNetV2 model. Further, the extracted features are provided as input to a Bayesian hyperparameter-optimized Random Forest Classifier. The Random Forest Classifier categorizes frames into classes based on features and then sorts frames by entropy values within each class. However, outliers may occasionally contain valuable information. The proposed approach examines frames from each predicted class to avoid discarding crucial information owing to outliers.

The rest of the paper is structured as follows: Section II details the methodology employed for WCE video summarization. Section III presents the experimental results and provides a thorough analysis. The final section summarizes the main conclusions drawn from this study.

## II. RELATED WORK

There have been several studies that aim for video summarization in WCE. The approaches for video summarization can be categorized into two categories: The first is a Generic approach that primarily focuses on Keyframe extraction or Shot Segmentation, and the second is Disease identification specific. Several approaches use keyframe extraction techniques for video summarization. Keyframe extraction is a technique for video summarization that involves

extracting the most informative frames from a long video that has many redundant frames. Researchers have explored various criteria and algorithms for selecting keyframes representing essential video information.

### A. Generic Approaches

A general approach to video summarization in WCE is keyframe identification or Shot Segmentation. Keyframes are the frames that contain the most informative part of the video and can be considered as a representative frame of the entire video. On the other hand, Shot segmentation is dividing a long video into small shots. A shot is a part of the video that contains similar frames. A technique for keyframe extraction that uses depth maps not only for keyframe identification but also for localization of the polyps was proposed in [19]. The proposed approach detects the keyframes that contain the polyps. In addition to the depth information, the proposed technique utilizes the image moments and edge magnitudes to select the keyframes.

The author in [20] proposed a method for generating video summaries by utilizing the deep features extracted by using InceptionV3 and then using the K-Means clustering algorithm to group similar frames in one cluster. Frames from the clusters are selected to generate a final summary. A technique for keyframe extraction that first extracts the deep features of the frames using a Convolutional Autoencoder Neural Network (CANN) was used in [16]. The frames are then grouped into similar and dissimilar frames, representing shots of the WCE video. From each shot, keyframes are then selected using motion analysis.

### B. Disease-Specific Approaches

Specific approaches focus on disease identification along with video summarization. One of the most common signs of gastrointestinal abnormality is bleeding. Several researchers worked towards identifying bleeding regions, polyps, ulcers, erosions, and tumors in the WCE Videos.

1) *Bleeding detection*: The author in [10] proposed a method for automatically detecting and segmenting bleeding regions by leveraging a CNN structure. The CNN utilized by the author is of low complexity.

2) *Polyps*: Most polyp detection approaches utilize the shape or texture features for polyp identification. However, [4] used an approach that considers both the context and shape of the polyp. Using a context-based reduces the chances of misclassifying some other polyp-like structure as a polyp. Whereas shape features help to capture the geometric information of polyps.

In [5] the author used a synthetically designed high-dimensional feature using Local Binary Patterns – Local, Uniform, and Complete, combined with the Histogram of oriented gradients (HOG). The high-dimensional descriptors provide the visual words as output, after they are fed as an input of the K-means clustering method. Finally, SVM and Fisher's linear discriminated analysis (FLDA) are used for polyp classification. The author in [6] combined the textural features and the Local Fractal Dimensions. The proposed method first detects the keypoints of the frames using SIFT



followed by the textural feature extraction of the neighborhood of the keypoints. In the end, the classification is carried out by an SVM classifier.

3) *Ulcers*: A two-staged automated ulcer detection system was proposed in [9]. In the first stage, superpixels are identified. A superpixel is a group of pixels under some restriction of local image features such as color, intensity, or texture. Then, the saliency regions are identified based on texture and color. The color and texture-based saliency maps are fused together to create a better salient representation of the ulcers. In the second stage, ulcer classification is done using a Bag of Words Model. A CNN-based approach to detect small intestinal ulcers and erosion in WCE images was proposed in [12].

4) *Crohn's disease*: The authors in [7] developed a CNN (Convolutional Neural Network) to classify WCE images into two categories- normal images and the images that are likely to have evidence of Crohn's lesions.

5) *Tumor*: A tumor is a mass of abnormal cells. They can be cancerous in rare situations. Researchers are exploring this field to develop techniques for automated detection of tumors in the WCE images. The author in [8] exploit the image's color and textural features; later, a support vector machine (SVM) is utilized for feature selections for tumor detection.

Considering the different approaches for WCE video summarization it can be concluded that a generic approach that is able to identify the abnormal frames of a WCE video is a better approach for video summarization. An approach that identifies only a specific disease cannot find other diseases, there may be a case where a patient has multiple abnormalities. Therefore a generic method for WCE video summarization is a better approach for generating WCE video summaries. The video summaries generated by a generic method can further be evaluated by the gastroenterologist for pinpointing the particular ailment.

### III. METHODOLOGY

The proposed method leverages the benefits of classification for video summarization. It generates a video summary of a long video. A video is an ordered set of frames. For processing a video, the first step is to generate the frames from the videos. CEV is the Capsule Endoscopy Video, which can be represented as

$$CEV = \{f_1, f_2, f_3, \dots, f_n\}$$

Where,  $\{f_1, f_2, f_3, \dots, f_n\}$  represents the ordered set of frames. Video summarization is the process of extracting the informative frames of the video. Let VS be the Video summary

then,  $VS \subset CEV$ , and  $VS = \{f_1, f_2, f_3, f_4, \dots, f_k\}$  where  $k < n$ . We may assume this by renumbering the frames.

Fig. 1 depicts a block diagram of the proposed video summarization approach.

#### Algorithm 1

Step 1: Generating frames from the video.

$CEV = \{f_1, f_2, f_3, \dots, f_n\}$ , where  $f_1, f_2, f_3, \dots, f_n$  are the video frames.

Step 2: Extract features from the frames by using pre-trained MobileNetV2.

For  $i=1$  to  $n$  do

$y_i \leftarrow M(f_i)$ ,  $M(f_i)$  represents the MobileNetV2 used for feature extraction

Projecting each feature vector  $y_i$  to embedding  $\gamma$

$\gamma_i \leftarrow y_i$

$\gamma = \{\gamma_1, \gamma_2, \dots, \gamma_n\}$

Step 3: The Random Forest algorithm processes the input as feature vectors extracted from frames in Step 2.

Step 4: Identifying the best parameters for the Random Forest using Bayesian Optimization.

Step 5: Testing the Model and generating Video Summaries.

#### C. MobileNetV2

MobileNet architecture was introduced by Google in 2018. As the name indicates, MobileNet is a lightweight convolutional neural network developed for mobile or embedded devices. It works efficiently with devices that have limited computational resources. MobileNetV2's first layer is a depth-wise convolution that performs lightweight filtering by applying a single convolutional filter per input channel. The second layer is point-wise convolution, which computes linear combinations of the input channels to generate new features. The fewer parameters and matrix multiplications significantly contribute to reducing the complexity of MobileNetV2. Some key features of MobileNetV2 are:

- The depth-wise convolution layer applies a separate filter to each input channel, producing intermediate feature maps. A point-wise convolution followed them to combine these features linearly. This two-step convolution approach significantly reduces the number of parameters and computations compared to traditional convolutional layers.
- Inverted residual blocks of MobileNetv2 are capable of capturing more complex patterns. An inverted residual block consists of a bottleneck layer, which downsizes the number of input channels followed by a depth-wise separable convolution and a linear projection layer to upsize the number of channels back. Skip connections are also used to retain low-level features.

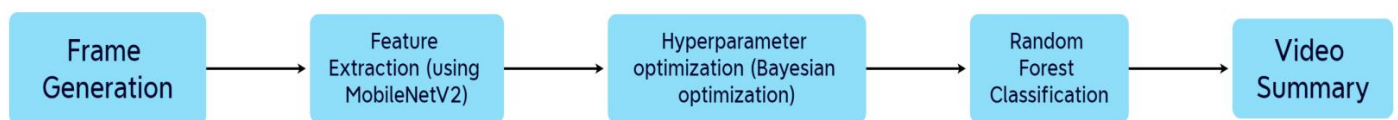


Fig. 1. Block diagram of video summarization.

- There are two hyperparameters in MobileNetV2- width multiplier and resolution multiplier. The width multiplier reduces the number of channels in each layer. The resolution multiplier scales down the input image size. Due to the resolution multiplier, some spatial information is sacrificed. The hyperparameters reduce the model's complexity and the computational requirements.

#### D. Random Forest

Random Forest [22] is a supervised machine-learning algorithm developed by Leo Breiman. It uses an ensemble of multiple decision trees for generating predictions. Ensemble means combining multiple models. Thus, a Random Forest uses a collection of decision trees to make predictions rather than an individual decision tree. The random forest algorithm delivers a precise and cohesive output by aggregating these tree's outputs. Fig. 2 shows the working of a random forest classifier. The Random Forest classifier classifies an image in one out of the different output classes. Every decision tree casts a vote to which the input vector belongs. However, in the WCE video, multiple frames need to be classified; there may be a part of a video that contains redundant information, and a small part of the video contains abnormality. In other words, the outlier may contain the most informative information. To avoid missing any informative part of the endoscopic video, a technique for video summarization is adopted that incorporates a change in the voting module of the random forest. The proposed voting module calculates the entropy of each image. The total number of votes an output class received is sorted based on the entropy values, and the top 10% frames from each class are combined to generate the video summary.

#### Algorithm 2

*Step 1: In the Random forest model a subset of features are randomly selected and decision trees are constructed from each sample.*

for  $t=1$  to  $T$  do

$D = \{D_1, D_2, D_3, \dots, D_T\}$ , where

$D$  is the set of decision trees

$T$  is the total number of decision trees

$D_t \subset \gamma$ , where  $1 < t < T$

*Each decision tree is constructed from a subset of feature embedding  $\gamma$ .*

*Step 2: For each input image, each decision tree will generate an output and cast a vote to one of the output classes.*

If  $L = \{L_1, L_2, L_3, \dots, L_k\}$  and  $C = \{C_1, C_2, C_3, \dots, C_k\}$  represents the set of labels of the Output class and  $C$  represents the count for each class label then,

for  $i=1$  to  $n$  do

for  $t=1$  to  $T$  do

$P_{it} \leftarrow D_t(y_i)$ ,  $P$  is the predicted Label for  $y_i$  and  $P_{it}$  it takes value from  $L$ , and update the corresponding Label's count

*Step 3: Final output is considered based on Majority Voting for Classification. However, for every  $L_i$ , where  $1 < i < k$  The proposed voting module also calculates the entropy-based ranks for each vote cast for each class.*

*Step 4: The top 10% of the total votes (frames) that a class got are selected to generate a final video summary. And the final summary VS is generated for the video VCE.*

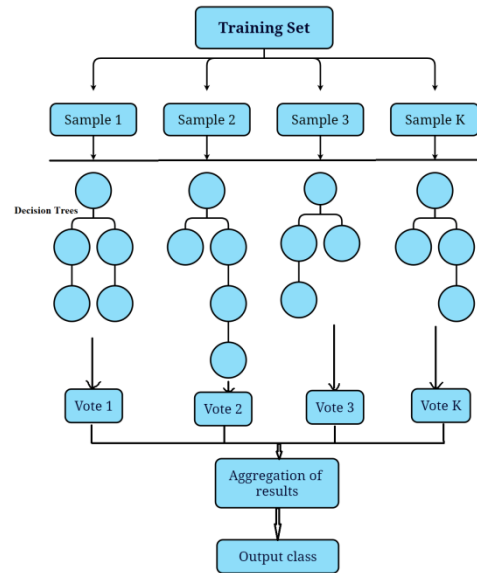


Fig. 2. Working of random forest classifier.

Each decision tree of the Random forest is constructed by using a different random sample from the training data every time, reducing the chances of overfitting. This tree construction task can be run on multiple CPU cores, reducing the training time. Random forests' voting and aggregation feature helps them effectively deal with missing and noisy data. A Random forest's most significant advantage is dealing with small sample sizes, high-dimensional feature space, and complex data structures.

#### E. Hyperparameter Tuning

Hyperparameter tuning is finding a hyperparameter setting for a machine-learning model to increase its accuracy. There are several techniques of hyperparameter optimization. Grid Search and Random Search are the most reliable techniques for hyperparameter tuning. Grid Search is an exhaustive technique; it considers each possible combination of hyperparameters to determine the optimum value. However, Random Search explores random values of the hyperparameters in a given search space. Both techniques are not adaptable; the results generated every time are independent of the previous outcomes. Random and Grid searches are significantly slow because of their exhaustive nature for search space exploration. Bayesian optimization [23] uses a probabilistic model to guide the search, which helps explore the hyperparameter space more intelligently and reduces the number of evaluations needed, making it one of the computationally efficient techniques. The optimized values of hyperparameters after Bayesian optimization are depicted in Table I.

#### IV. EXPERIMENTS AND RESULTS

The experiments were implemented in Python, and the training and validation dataset was obtained from Kaggle

(WCE Curated Colon Disease Dataset Deep Learning) [24]. This data set consists of images a wireless capsule captures during endoscopy to diagnose abnormal conditions. It has three sets: training set, test set, and validation set. The WCE dataset has labeled images. The dataset has four labels: Normal, Ulcerative Colitis, Polyps, and Esophagitis.

TABLE I. HYPERPARAMETER OPTIMIZATION RESULTS

Hyperparameter	Optimized Value	Significance
n_estimators	174	Number of trees in the forest
min_samples_split	2	Minimum number of samples required to split an internal node.
max_depth	15	Maximum depth of the tree
random_state	42	Controlling the randomness

The performance of the Random Forest is presented in the confusion matrix of Fig. 3. The predicted labels are close to the true labels of the images. It represents a model with high accuracy. The model obtained an accuracy of 98.75% over 50 iterations.

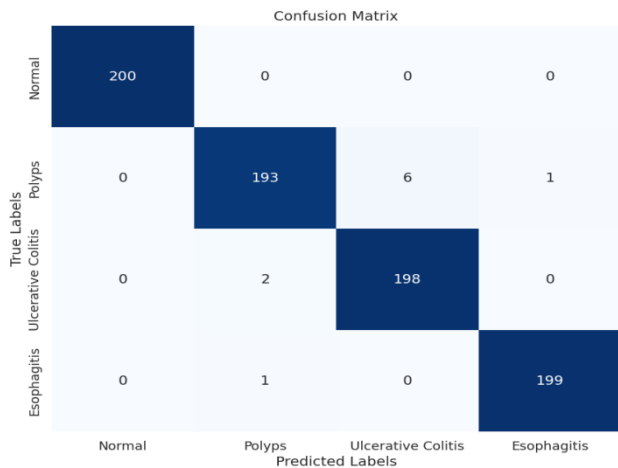


Fig. 3. Confusion matrix.

The model's performance is also compared over Precision, Recall, and F-score (Table II). F-Score is computed using (1). It is computed based on recall (2) and precision (3). Recall measures the proportion of actual positive instances that were correctly predicted as positive by the model.

$$F\_Score = \frac{(2 * Recall * Precision)}{(Recall + Precision)} \quad (1)$$

$$Precision = \frac{TP}{(TP + FP)} \quad (2)$$

$$Recall = \frac{TP}{(TP + FN)} \quad (3)$$

TABLE II. CLASSIFICATION REPORT

	Precision	Recall	F-Score
Normal	1.00	1.00	1.00
Polyps	0.97	0.97	0.97
Ulcerative Colitis	0.97	0.97	0.97
Esophagitis	1.00	1.00	1.00

True Positives (TP) are the positives that are correctly predicted as positives. False Positives (FP) are the negatives incorrectly predicted as positives. True Negatives (TN) are the negatives that are correctly predicted as negatives. False Negatives (FN) are the positives that are incorrectly predicted as negatives.

The average value of the F-score is obtained as 0.985. The classification report indicates that the model is 100% precise in predicting the Normal and Esophagitis labels.

The proposed voting module of the Random Forest Classifier not only votes for a particular class but also calculates the entropy of each image. Fig. 4 shows the predicted frames of the output classes: Polyps, Ulcerative Colitis, and Esophagitis according to the decreasing entropy values. The Final summary is generated by combining the top 10% of the frames from every predicted class, as shown in Fig. 5. The selection of 10% of the total images of a particular class was experimentally determined. The contribution of each output class to generate the final summary ensures that outliers don't get missed.

Fig. 5 shows a video summary of a patient suffering from Esophagitis. Although the disease identified is esophagitis the video summary generated has an abnormal bleeding part of the esophagus, which may otherwise have been missed.

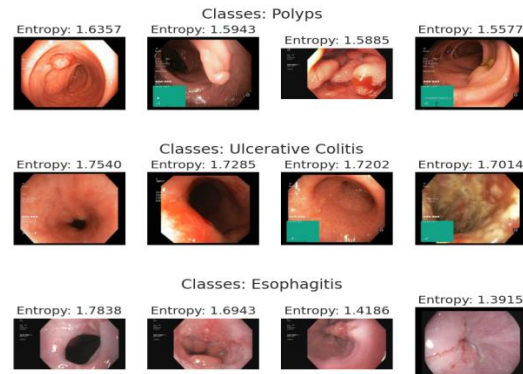


Fig. 4. Images from different output classes in decreasing order of entropies.



Fig. 5. Video Summary generated for an esophagitis patient.

## V. CONCLUSION

This paper introduces a WCE Video Summarization technique that uses transfer learning, random forest, and an entropy-based ranking mechanism to select informative frames and generate the video summary. Using MobileNetv2 for feature extraction allowed for prompt and efficient results to be obtained with excellent computational efficiency. Moreover, employing a Random Forest reduces the chances of overfitting. Selecting the most informative frames from each predicted class prevents the exclusion of outliers with valuable content. The proposed approach obtained an accuracy of 98.75% in

classifying the disease, and the Video summary generated by the model has a significantly reduced viewing time. In the future, a scalable WCE video summarization technique can be proposed that predicts the disease and maintains the temporal relation of the frames.

#### REFERENCES

- [1] G. J. Iddan, G. Meron, A. Glukhovskiy, and P. Swain, "Wireless capsule endoscopy," *Nature*, vol. 405, no. 6785, p. 417, May 2000, doi: 10.1038/35013140.
- [2] P. Swain, "Wireless capsule endoscopy," *Gut*, vol. 52, no. 90004, pp. 48iv–4850, Jun. 2003, doi: 10.1136/gut.52.suppl\_4.iv48.
- [3] A. Wang et al., "Wireless capsule endoscopy," *Gastrointestinal Endoscopy*, vol. 78, no. 6, pp. 805–815, Dec. 2013, doi: 10.1016/j.gie.2013.06.026.
- [4] N. Tajbaksh, S. R. Gurudu and J. Liang, "Automated Polyp Detection in Colonoscopy Videos Using Shape and Context Information," in *IEEE Transactions on Medical Imaging*, vol. 35, no. 2, pp. 630–644, Feb. 2016, doi: 10.1109/TMI.2015.2487997.
- [5] Y. Yuan, B. Li and M. Q. -H. Meng, "Improved Bag of Feature for Automatic Polyp Detection in Wireless Capsule Endoscopy Images," in *IEEE Transactions on Automation Science and Engineering*, vol. 13, no. 2, pp. 529–535, April 2016, doi: 10.1109/TASE.2015.2395429.
- [6] M. E. Ansari and S. Charfi, "Computer-aided system for Polyp detection in wireless capsule endoscopy images," 2017 International Conference on Wireless Networks and Mobile Communications (WINCOM), Rabat, Morocco, 2017, pp. 1–6, doi: 10.1109/WINCOM.2017.8238211.
- [7] D. Marin-Santos, J. A. Contreras-Fernandez, I. Perez-Borrero, H. Pallares-Manrique, and M. E. Gegundez-Arias, "Automatic detection of Crohn disease in Wireless Capsule Endoscopic images using a deep convolutional neural network," *Applied Intelligence*, vol. 53, no. 10, pp. 12632–12646, Sep. 2022, doi: 10.1007/s10489-022-04146-3.
- [8] B. Li and M. Q. -h. Meng, "Tumor Recognition in Wireless Capsule Endoscopy Images Using Textural Features and SVM-Based Feature Selection," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 3, pp. 323–329, May 2012, doi: 10.1109/titb.2012.2185807.
- [9] Y. Yuan, J. Wang, B. Li, and M. Q. -h. Meng, "Saliency Based Ulcer Detection for Wireless Capsule Endoscopy Diagnosis," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 2046–2057, Oct. 2015, doi: 10.1109/tmi.2015.2418534.
- [10] M. Hajabdollahi, R. Esfandiarpour, K. Najarian, N. Karimi, S. Samavi and S. M. Reza Soroushmehr, "Low Complexity CNN Structure for Automatic Bleeding Zone Detection in Wireless Capsule Endoscopy Imaging," 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 2019, pp. 7227–7230, doi: 10.1109/EMBC.2019.8857751.
- [11] O. Bchir, M. M. B. Ismail, and N. Alzahrani, "Multiple bleeding detection in wireless capsule endoscopy," *Signal, Image and Video Processing*, vol. 13, no. 1, pp. 121–126, Jul. 2018, doi: 10.1007/s11760-018-1336-3.
- [12] M. L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Jan. 2001, doi: 10.1023/a:1010933404324.
- [13] S. Fan, L. Xu, Y. Fan, K. Wei, and L. Li, "Computer-aided detection of small intestinal ulcer and erosion in wireless capsule endoscopy images," *Physics in Medicine & Biology*, vol. 63, no. 16, p. 165001, Aug. 2018, doi: 10.1088/1361-6560/aad51c
- [14] A. Caroppo, P. Siciliano, and A. Leone, "An expert system for lesion detection in wireless capsule endoscopy using transfer learning," *Procedia Computer Science*, vol. 219, pp. 1136–1144, Jan. 2023, doi: 10.1016/j.procs.2023.01.394.
- [15] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770–778, doi: 10.1109/CVPR.2016.90.
- [16] A. Biniaz, R. A. Zoroofi and M. R. Sohrabi, "Automatic reduction of wireless capsule endoscopy reviewing time based on factorization analysis," *Biomedical Signal Processing and Control*, vol.59, p.101897, May 2020, doi:10.1016/j.bspc.2020.101897.
- [17] B. Sushma and P. Aparna, "Summarization of Wireless Capsule Endoscopy Video Using Deep Feature Matching and Motion Analysis," in *IEEE Access*, vol. 9, pp. 13691–13703, 2021, doi: 10.1109/ACCESS.2020.3044759.
- [18] S. Adewole et al., "Unsupervised shot boundary detection for temporal segmentation of long capsule endoscopy videos.," arXiv (Cornell University), Oct. 2021, [Online]. Available: <http://export.arxiv.org/abs/2110.09067>
- [19] S. Adewole et al., "Graph Convolutional Neural Network For Weakly Supervised Abnormality Localization In Long Capsule Endoscopy Videos," 2021 IEEE International Conference on Big Data (Big Data), Dec. 2021, doi: 10.1109/bigdata52589.2021.9671281.
- [20] P. Sasmal, A. Paul, M. K. Bhuyan, Y. Iwahori, and K. Kasugai, "Extraction of Keyframes From Endoscopic Videos by using Depth Information," *IEEE Access*, vol. 9, pp. 153004–153011, Jan. 2021, doi: 10.1109/access.2021.3126835.
- [21] V. G. Raut and R. Gunjan, "Transfer learning based video summarization in wireless capsule endoscopy," *International Journal of Information Technology*, vol. 14, no. 4, pp. 2183–2190, Feb. 2022, doi: 10.1007/s41870-022-00894-0.
- [22] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, 2018, pp. 4510–4520, doi: 10.1109/CVPR.2018.00474
- [23] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Jan. 2001, doi: 10.1023/a:1010933404324.
- [24] F. Hutter, H. H. Hoos, and K. Leyton-Brown, "Sequential Model-Based optimization for general algorithm configuration," in *Springer eBooks*, 2011, pp. 507–523. doi: 10.1007/978-3-642-25566-3\_40.
- [25] "WCE Curated Colon Disease Dataset Deep Learning," Kaggle, Apr. 15, 2022. <https://www.kaggle.com/datasets/francimon/curated-colon-dataset-for-deep-learning>.

# Contributed Factors in Predicting Market Values of Loaned Out Players of English Premier League Clubs

Muhammad Daffa Arviano Putra<sup>1</sup>, Deshinta Arrova Dewi<sup>2</sup>,  
Wahyuningdiah Trisari Putri<sup>3</sup>, Retno Hendrowati<sup>4</sup>, Tri Basuki Kurniawan<sup>5</sup>

Department of Informatics-Faculty of Engineering Science, Paramadina University, Jakarta, Indonesia<sup>1,3,4</sup>  
Faculty of Data Science and Information Technology-INTI International University, Nilai, Malaysia<sup>1,2</sup>  
Postgraduate Program of Information Technology-Bina Darma University, Palembang, Indonesia<sup>5</sup>

**Abstract**—The top tier of the English football league division is occupied by the English Premier League (EPL). It has become a global phenomenon with exhilarating skills and has been one of the most-watched professional football leagues on the planet. The possibility of a player temporarily playing for a club other than the one to whom they are now contracted is known as a "loan player" in the English Premier League (EPL) hence, each player has a market value. Market value is an estimate of how much a player costs when a club wants to buy his contract from another club. The purpose of this study is to determine the factors that influence a player's market value at the conclusion of a loan period. With the Transfermarkt player transfer record dataset for the years 2004 through 2020, we use linear regression analysis. Our study found that a football player's market worth at the end of a loan period is influenced by several aspects, including market value at the beginning, goals, appearances, and total loan.

**Keywords**—Data analytics; predicting market value; English Premier League; loaned out players; consumption; resource use

## I. INTRODUCTION

Football is one of the most popular team sports worldwide [1]. According to research conducted by Nielsen Sports, more than 40% of people aged 16 or older in countries with high populations and large markets said they are "interested" or "very interested" in following football [2]. The most prominent and well-known football league in the world is the English Premier League (EPL). The EPL has a lot of fans all over the world. According to the official website of the English Premier League, the cumulative global audience of EPL for season 2018/2019 was over 3 billion [3]. Due to the huge fans and excitement of the EPL, it has succeeded in attracting the interest of investors. Matchday revenue, broadcast deals, and commercial activity are some of the ways a club can generate a lot of profit [4].

To maximize revenue, the club must be popular among the viewers and have winning matches. Therefore, every owner strives to strengthen their club squad to be competitive and have a successful season. One way to strengthen a club is to buy good and talented players. Some of the examples we saw recently when Manchester City bought Jack Grealish from Aston Villa with a figure of €117.50m, Chelsea bought Romelu Lukaku from Inter Milan for €115.00m, and Manchester United bought Jadon Sancho for €85.00m from the German club, Borussia Dortmund [5].

Each player has a market value. Market value is an estimate of how much a player costs when a club wants to buy his contract from another club [1]. The price does not apply if a club only loans a player. If a club wants to loan a player, they are most likely only required to pay the player's salary for the duration of the loan. However, the player will still carry a market value that can go up or down when playing at the loan club. During the loan period, the player might perform extraordinarily and make many appearances and therefore, his market value will increase and vice versa. In every transfer window, every club in the EPL is likely to loan out their players to make room for their squad or give players a chance to build a reputation at another club. For this study, we use the market value published by the Transfermarkt website. The market values provided by the website are economically relevant and are viewed as having a fine reputation in the sports industry [6]. A study by Peeters revealed that Transfermarkt crowd valuation is referenced privately by club officials during player contract negotiations because it is more accurate than other valuations such as FIFA ranking and the ELO rating [7].

The academic community has found the football transfer market to be an engaging subject [8]. Additionally, we believe that additional investigation into the market value of football players would be an intriguing topic to pursue. Fortunately for us, the information required to do so is readily accessible through websites devoted to the sport of football. The general contributing aspects to a player's market value are not often covered in academic publications, nevertheless. Since the market value at the conclusion of the loaned time in EPL is greater than average, we are looking for contributing causes for that higher market value in this study.

### A. Data Availability Statement

The data collected for the model is from Transfermarkt, a German Website that provides football information and data, such as scores, league tables, club squads, and many more using web scraping. Football-related research has used this website as its source data. The website incorporates crowdsourcing to estimate a player's market value in several professional football leagues. This means that every person can join the community and discuss the market value of any football player.

## II. PREVIOUS STUDIES

There were studies on the subject of prediction of the market value of a football player that aligned with this research. Such as one from Singh and Lamba that suggested consistency, popularity, crowd estimation, and performance parameters enhance the prediction accuracy of the market value of football players [9]. While Felipe et. al. stated in their paper that the playing position (attacking midfielders) and age of the player (born in the first quarter of the year) are the most economically valued in terms of current value and maximal value [10]. Müller, Simons, and Weinmann stated that there are three categories of indicators of the market value of a player. There are the player characteristics (age, height, position, footedness, nationality), player performance (playing time, goals, assists, passing, dribbling, dueling, fouls, and cards), and player popularity (news, internet links) [1]. Further study on the popularity component, Frenger et. al. suggested that social media activities significantly influenced a football player's market value on the site [11].

A study on the player performance is also done by Richau et. al. which emphasized the actual performance of a football player to determine their market value, the paper stated that the actual performance is measured through individual player's age, minutes played, offense, defense, and team and analyzed using boosted regression trees [12]. They found that individual player performance indicator does not have the highest influence on the market value; instead, they found that team dimension average rank influences market value. Along this line, Metelski tried to find factors affecting the value of football players in the transfer market for the Polish Football League using descriptive statistics and several statistical tests. The writer uses the position on the pitch, age at transfer, year of transfer, the destination country, and selling club as the indicators. They found that the age of the player is a significant factor in the football players' values [13]. Moreover, Behravan and Razavi used the FIFA 20 dataset of various performance ratings of 18,278 players. Their novelty is the use of an automatic clustering algorithm in the first phase which they called APSO-clustering, and further training of a hybrid regression method called PSO-SVR for each cluster [14] that can estimate the players' value with an accuracy of 74%.

## III. METHOD

Fig. 1 illustrates the methods we use for this study. These steps will be more specifically described afterward.

### B. Data Collection

The data collected for the model is from Transfermarkt, a German Website that provides many footballs information and data, such as scores, league tables, club squads, and many more using web scraping. Football-related research has used this website as their source data, such as in [14] [15][16][6][17]. The website incorporates crowdsourcing to estimate a player's market value in several professional football leagues. This means that every person can join the community and discuss the market value of any football player. Everyone can suggest a market value for a player with good arguments and reasons to justify the player's estimation [6]. However, not everyone's opinion has the same value. The Transfermarkt website has

data on past loan players' performance in the EPL from the season 2004/2005 to 2021/2022. However, as the EPL 2021/2022 season is still ongoing at the time of the writing of this paper, we will limit the data to season 2020/2021.

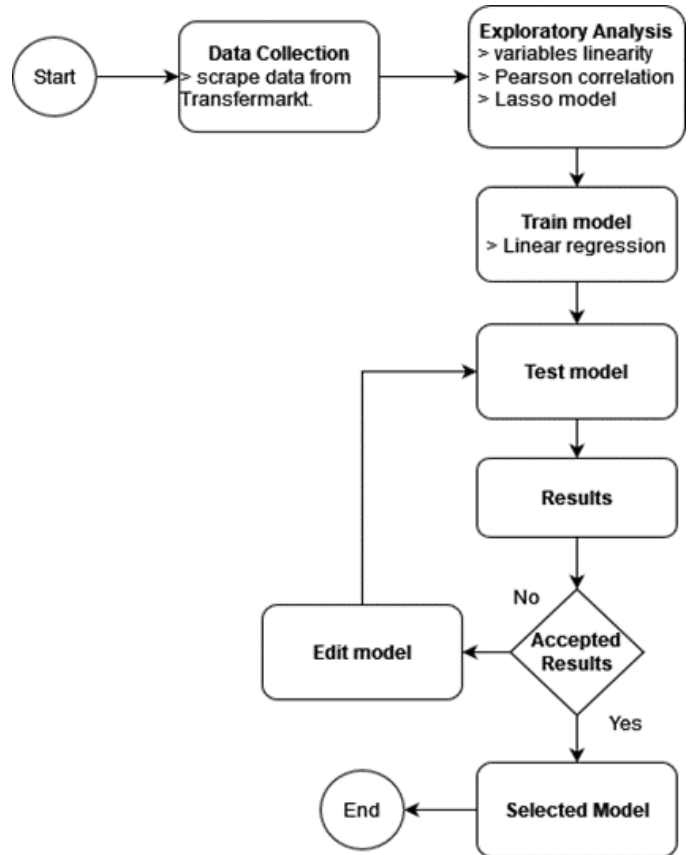


Fig. 1. Flowcharts of an applied method for the research.

There are 10 variables in the data as captured in Table I. In this study, we want to predict the last variable, which is the total market value of the loaned-out players from a club in a season. Therefore, the first nine variables are labeled as the explanatory variables and the last variable is our target or dependent variable.

### C. Exploratory Analysis

In this section, we explore the data that we have. Our data has 339 rows and 10 columns, which means we have a total record of 339 data of all clubs who loaned out their players from season 2004/2005 to 2020/2021. Every season consists of 20 clubs and because the EPL uses a promotion and relegation system, there can be more than 20 unique clubs in the data. The first thing we have done is to list and count every unique club in the data. The outcome is that we have 40 unique clubs that will be trained in the Linear Regression model. We start exploring our data with the perspective of the relationship between variables, the data distribution, the correlation among variables, Least Absolute Shrinkage and Selection Operator (LASSO) in identifying features that may be the contributing factors to predicting market values of the loaned players.

TABLE I. SUMMARY OF THE AFM INFORMATION OF CDS QDS

Variable Name	Description
name	The club's name
year	The year of loaned out players data of a club (ranging from 2004 to 2020)
total_loan	The total loaned-out players from the club in the respective year
average_loan (in years)	The average number of loan periods of all loaned out players from the club in the respective year
appearances	The total number of appearances from all loaned-out players from the club in the respective year
starting_formation	The total number of appearances in the starting formation from all loaned-out players from the club in the respective year
goals	The total number of goals from all loaned-out players from the club in the respective year
average_minutes_played	The average minutes played by all loaned-out players from the club in the respective year (the maximum is 90 as a football match is played for 90 minutes)
market_value_at_start (in M €)	The total market values at the start of the loan period from all loaned-out players from the club in the respective year
market_value_at_end (in M €)	The total market values at the end of the loan period from all loaned-out players from the club in the respective year

D. Checking the Relationship between Variables using Scatter Plots

We start by checking the bivariate relationships between eight variables of our data. The total loan personnel, average loan time in years, number of appearances made by the players, starting formation of the players, goals made, average minutes played, market value at the start, and the last variable is the market value at the end. As we can see from the scatter plot in Fig. 2, there are three types of relationships shown, discrete relationships, random, and linear.

The discrete plot was obtained from the total loaned variables as there were only two values in these variables as the players were loaned only for 1 year or 2 years. The random relationship is shown by the total loaned players' variable against average minutes played, market value at the start, and market value at the end. The random relationship we see with average minutes played against all other variables. While the rest of the bivariate combinations showed some kind of linear relationship.

In this paper, we focused on the relationship between the market value at the end with the rest of the variables, so we found that the market value at the end has a strong linear relationship with the market value at the start. We explore the relationship more in the sections below.

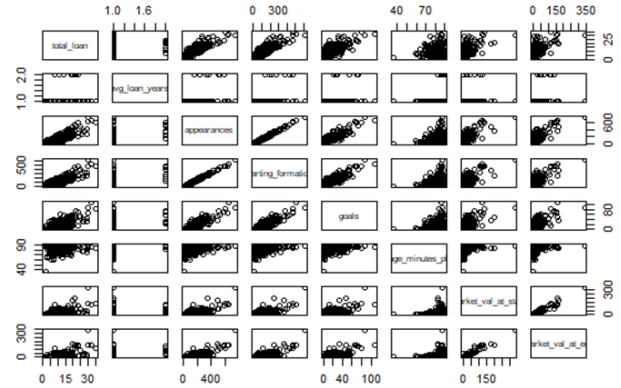


Fig. 2. Bivariate relationship of variables data.

E. Checking the Univariate Distribution using Histogram

To see if the observed data represent a random sample from the population; we use the histogram to check the distribution. Fig. 3 below shows the distribution of seven variables that we consider from the data; we did not include the name and year variables because they are categorical. The initial histogram showed left and right skewed data, so we do a log transformation on the variables to stabilize the variance, such as seen in [18] and [19]. Ensuring the log transformation, the average minutes played variable is still left skewed, while the average loan years is discrete.

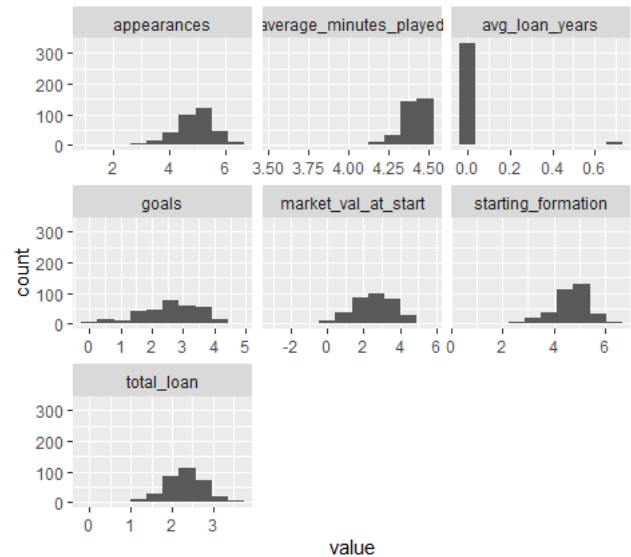


Fig. 3. Log transformed data distribution of market value predictors.

F. Pearson Correlation

The next thing we do in our exploratory step is to see the Pearson correlation to determine the precise extent or degree of any connection between any two variables, indicating its presence or absence.







that model 1 has the least R2 values, at an average of 88.9%, 87.9%, and 87.2% for respective 3, 5, and 10 folds. The percentage showed how well the dependent variable, market value at the end can be accounted for by the nine predictors. Model 2 came second with the average R2 values of 89.6%, 88.7%, and 88.3% accounted for by seven predictors, and model 3 has the highest average R2 values of 90.5%, 90.3%, and 90.1% accounted for by five predictors.

In terms of accuracy, the mean RMSE for model 1 for 3, 5, and 10 folds, respectively are 8.69525, 7.91934, and 7.24353. For Model 2, the mean RMSE are 8.10107, 7.51161, and 6.84796, respectively. For Model 3, the mean RMSE are 0.33339, 0.32948, 0.31620, respectively. Model 3 accuracy is higher than the two previous models. Therefore, we can conclude that removing the two variables average minutes played, and average loan years, is the right decision. With model 3 we come up with this linear equation:

$$\begin{aligned} \text{market\_val\_at\_end} = & (-2.282 - 0.04653 * \text{total\_loan}) + (0.5237 * \\ & \text{appearances}) - (0.00007325 * \text{starting\_formation}) + (0.1056 * \\ & \text{goals}) + (0.8716 \text{ market\_val\_at\_start}) \end{aligned} \quad (1)$$

With the use of the aforementioned equation, we can observe that while every variable affects the market value at the end, the market value at the beginning, objectives, appearances and total loan all has positive correlations and significant contributions. This differs slightly from our initial hypotheses based on the Pearson correlation, which was initial market value, early appearances, and initial formation.

The majority of the contributing factors to the loan player in the ELP have generally been identified by our investigation. With this investigation and its outcome, we have discovered a previously unknown association. We have identified which variables are connected to or have the strongest relationships with, and we may be able to identify patterns within the dataset as a result of this understanding.

## V. CONCLUSION

In this study, we identified the elements that contributed to a football player's market worth in the English Premier League at the end of the loan period. Market value at launch, goals, appearances and total loan is among them. The market worth of a football player after a loan has been made is something we can forecast using exploratory research and a linear regression model. To discover the best predictor, we tested three distinct models. Our study revealed that the predictors differed slightly from what we had initially thought. Although linear regression is simple to comprehend and explain, we believe the model is adequate for use in this investigation. In future experiments, we hope to incorporate more data into our model, as the current data is very limited. We can also implement other models to better understand the contributing factors of a football player's market value.

## REFERENCES

- [1] O. Müller, A. Simons and M. Weinmann, "Beyond crowd judgments: Data-driven estimation of market value in association football," *European Journal of Operational Research*, vol. 263, no. 2, pp. 611-624, 2017.
- [2] "World Football Report 2018," Nielsen Sports, 2018. [Online]. Available at: <https://www.nielsen.com/wp-content/uploads/sites/3/2019/04/world-football-report-2018.pdf>.
- [3] "Premiere League Global Audience on The Rise," Premier League, 2019. [Online]. Available: <https://www.premierleague.com/news/1280062>.
- [4] T. Dima, "The Business Model of European Football Club Competitions," *Procedia Economics and Finance*, vol. 23, no. Oct. 2014, pp. 1245-1252, 2015.
- [5] "Premiere League - Transfer records," Transfermarkt, 2021. [Online]. Available: [https://www.transfermarkt.com/premier-league/transferrekorde/wettbewerb/GB1/plus/galerie/0?saison\\_id=2021&land\\_id=alle&ausrichtung=&spielerposition\\_id=alle&altersklasse=&leih=&w\\_s=s&zuab=zu](https://www.transfermarkt.com/premier-league/transferrekorde/wettbewerb/GB1/plus/galerie/0?saison_id=2021&land_id=alle&ausrichtung=&spielerposition_id=alle&altersklasse=&leih=&w_s=s&zuab=zu).
- [6] S. Herm, H.-M. Callsen-Bracker and H. Kreis, "When the crowd evaluates soccer players' market values: Accuracy and evaluation attributes of an online community," *Sport Management Review*, vol. 17, no. 4, pp. 484-492, 2013.
- [7] T. Peeters, "Testing the Wisdom of Crowds in the field: Transfermarkt valuations and international soccer result," *International Journal Forecasting*, vol. 34, no. 1, pp. 17-29, 2018.
- [8] D. Matesanz, F. Holzmayer, B. Torgler, S. L. Schmidt and G. J. Ortega, "Transfer market activities and sportive performance in European first football leagues: A dynamic network approach," *PLOS ONE*, vol. 13, no. 12, pp. 1-16, 2018.
- [9] P. Singh and P. S. Lamba, "Influence of crowdsourcing, popularity and previous year statistics in market value estimation of football players," *Journal of Discrete Mathematical Sciences and Cryptography*, vol. 22, no. 2, pp. 113-126, 2019.
- [10] J. L. Felipe, A. Fernandez-Luna, P. Burillo, L. E. de la Riva, J. Sanchez-Sanchez and J. Garcia-Unanue, "Money Talks: Team Variables and player Positions that Most Influence the Market Value of Professional Male Footballers in Europe," *Sports Policy and Finance*, vol. 12, no. 9, pp. 10-17, 2020.
- [11] M. Frenger, F. Follert, L. Richau and E. Emrich, "Follow me... on the relationship between social media activities and market values in the German Bundesliga," *Saarbrücken*, 2019. [Online]. Available: <http://www.soziooekonomie.org>.
- [12] L. Richau, F. Follert, M. Frener and E. Emrich, "Performance indicators in football: The importance of actual performance for the market value of football players," *SCIAMUS Sport and Manag*, vol. 4, pp. 41-61, 2019.
- [13] A. Metelski, "Factors affecting the value of football players in the transfer market," *Journal of Physical Education and Sport*, vol. 21, no. 2, pp. 1150-115, 2021.
- [14] I. Behravan and S. M. Razavi, "A novel machine learning method for estimating football players' value in the transfer market," *Soft Computing*, vol. 25, no. 3, pp. 2499-2511, 2021.
- [15] H. Adiwiyana, H. I. Adiwiyana and Harywaman, "Factors that Determine the Market Value of Professional Football Players in Indonesia," *J. Din. Akunt*, vol. 13, no. 1, pp. 51-61, 2021.
- [16] R. Stanojevic and L. Gyarmati, "Towards Data-Driven Football Player Assessment," *IEE Int. Conf. Data Min. Work. ICDMW*, vol. 0, pp. 167-172, 2016.
- [17] M. He, R. Cachucho and A. Knobbe, "Football Player's Performance and Market Value," in *Proc 2nd Work. Sport. Anal. Eur. Conf. Mach. Learn. Princ. Pract. Knowl. Discov. Databases (ECML PKDD)*, 2015.
- [18] D. Curran-Everett, "Explorations in statistics: the log transformation," *Adv. Physiol. Educ.*, vol. 42, no. 2, pp. 343-347, 2018.
- [19] H. Son, C. Hyun, D. Phan and H. J. Hwang, "Data analytic approach for bankruptcy prediction," *Expert Systems with Applications*, vol. 138, 2019.
- [20] Z. Yan and Y. Yao, "Variable selection method for fault isolation using least absolute shrinkage and selection operator (LASSO)," *Chemom. Intell. Lab. Syst.*, vol. 146, pp. 136-146, 2015.
- [21] S. Tian, Y. Yu and H. Guo, "Variable selection and corporate bankruptcy forecasts," *Journal of Banking & Finance*, vol. 52, no. December, pp. 89-100, 2015.

- [22] P. Ghosh, S. Azam, M. Jonkman and A. Karim, "Efficient Prediction of Cardiovascular Disease Using Machine Learning Algorithms with Relief and LASSO Feature Selection Techniques," *IEEE Access*, vol. 9, pp. 19304-19326, 2021.
- [23] M. R and R. R, "LASSO: A Feature Selection Technique in Predictive Modeling for Machine Learning," 2016 IEEE International Conference on Advances in Computer Application (ICACA), pp. 18-20, 2016.

# Exploring the Challenges and Impacts of Artificial Intelligence Implementation in Project Management: A Systematic Literature Review

Muhammad Irfan Hashfi, Teguh Raharjo

Faculty of Computer Science, Universitas Indonesia, Salemba, 10430, Indonesia

**Abstract**—This paper presents a systematic literature review (SLR) investigating the challenges and impacts of implementing artificial intelligence (AI) in project management, specifically mapping them into the process groups defined in the Project Management Body of Knowledge (PMBOK). The study aims to contribute to the understanding of integrating AI in project management and provides insights into the challenges and impacts within each process group. The SLR methodology was applied, and a total of 34 scientific articles were analyzed. The results and analysis reveal the specific challenges and impacts within each process group. In the Initiating Process Group, AI tools and analysis techniques address challenges in risk assessment, cost prediction, and decision-making. The Planning process group benefits from various tools and methodologies that improve risk assessment, project selection, cost estimation, resource allocation, and decision-making. The Execution process group emphasizes the importance of advanced tools and techniques in enhancing productivity, resource utilization, cost reduction, and decision-making. The Monitoring and Controlling process group demonstrates the potential of advanced tools in achieving efficiency, cost reduction, improved quality, and informed decision-making. Lastly, the Closing process group emphasizes the importance of utilizing advanced tools to minimize waste, optimize resource utilization, reduce costs, improve quality, and project closure success. Overall, this research provides valuable insights and strategies for organizations seeking to implement AI in project management, thereby enhancing the potential for success within the PMBOK Process Group.

**Keywords**—Artificial intelligence; project management; PMBOK process groups; challenge; impact

## I. INTRODUCTION

Industry 5.0 is an evolution of Industry 4.0, focusing on a more human-centric approach while leveraging advanced technologies like Artificial Intelligence (AI) and big data. It aims to create a sustainable and resilient industry by combining technological advancements with human needs [1], [2]. Project management has experienced a paradigm shift because of the quick development of AI technology. The way projects are planned, carried out, and controlled may be completely transformed using AI [3], [4]. The incorporation of AI into project management techniques, however, comes with a unique set of difficulties and has a big influence. Organizations need to address these challenges to fully harness the potential of AI in project management [5], [6]. Therefore, it is essential to investigate and understand the challenges faced and impacts

observed during the implementation of AI in project management [6].

Given the transformative impact of AI, there is an increasing significance in studying the process of AI adoption. Numerous studies have been conducted to explore the driving factors, barriers, challenges, and overall performance impact associated with AI implementation in organizations [6], [7]. The accuracy and appropriateness of data pertaining to project management tools are crucial. Initial AI tools for project management heavily depend on individuals to accurately input data, timely update tools, and make necessary corrections [8].

The integration of Artificial Intelligence (AI) into business operations has given rise to the emergence of intelligent environments, including advanced monitoring systems for project management. Similarly, AI has been embraced in the field of project management, offering promising prospects for the future of project management activities [3], [9]. Moreover, AI has the capability to monitor project status and make adjustments when required. Integrating AI into project management allows for minimal human intervention by utilizing intelligent machines and large data sets to automate decision-making and task management. This automation enables AI to play a guiding role in projects, automating tasks and aiding in decision-making processes [6], [10], [11].

This paper aims to conduct a systematic literature review (SLR) to investigate the challenges faced and impacts observed during the implementation of AI in project management, specifically mapping them into the process groups defined in the PMBOK [12], [13]. The research questions derived from this objective are as follows:

RQ1: What are the challenges faced and impacts observed during the implementation of AI in project management?

RQ2: How does the implementation of AI in project management incorporate the mapping of challenges and impacts into the process groups defined in the PMBOK?

By conducting a comprehensive review of existing literature, the findings will provide insights into the challenges and impacts specific to each process groups, helping organizations effectively address these challenges and leverage the potential benefits of AI in project management.

This paper follows a systematic process, commencing with a Literature Review section that evaluates existing studies on AI implementation and process groups in project management.

Then the Methodology section consists of three phases: planning the Systematic Literature Review (SLR), executing the SLR, and presenting the SLR findings. Within the Results and Analysis section, two subsections are delineated. The first subsection maps AI tools, implementation challenges, and impacts based on prior research, while the second subsection aligns AI tools, challenges, and impacts with PMBOK Process Groups (initiating, planning, execution, monitoring and controlling, closing). Lastly, the paper concludes by summarizing key findings, discussing their implications for project management, and suggesting directions for future research.

## II. LITERATURE REVIEW

### A. Artificial Intelligence and Data Mining

Artificial intelligence (AI) is a field within computer science that aims to develop intelligent machines for the benefit of humans. However, there is ongoing debate among researchers regarding the definition of AI due to its evolving nature. Various definitions exist in the literature, reflecting researchers' specializations and interests, all striving to explain the concept of AI and provide a wide range of technologies that enhance performance and interaction within organizations [3], [14]. AI has traditionally been approached from four perspectives: thinking and acting, which encompass thought processes, reasoning, and behavior. These perspectives can be further categorized into a human-centered approach based on human behavior observations, and a rationalist approach combining mathematics and engineering concepts [1].

Data mining is a branch of Machine Learning and Artificial Intelligence that involves analyzing a dataset using algorithms to uncover patterns and relationships. It allows users to analyze data from multiple perspectives, categorize information, and draw conclusions. The process involves searching for correlations between different fields in a database to extract valuable insights and information [15]–[17]. Machine learning enables the categorization of data processing methods in data mining into distinct categories such as classification, regression analysis, association rules, and clustering. Each of these mining methods can be executed using various machine learning techniques [18].

### B. Process Groups of Project Management

Project management procedures are organized into logical groups of inputs, tools and techniques, and outputs that are tailored to the needs of the organization, the project, and the stakeholders. Instead of being interchangeable with project phases, these process groups work together during each stage of the project's life cycle. In order to maintain flexibility and adaptation throughout the project, the number of iterations and interactions across processes can change based on the demands of the project [12], [13]. As shown in Fig. 1, participation with the remaining Process Groups is required to fully realize the collaborative aspect of project management.

Projects that adopt a process-based approach can be structured into five groupings based on different processes:

1) The initiating phase of a project involves defining and obtaining authorization for a new project or phase [12]. This

phase includes creating a project charter and implementing a formalized project initiation process to support project management decisions and ensure project success [19].

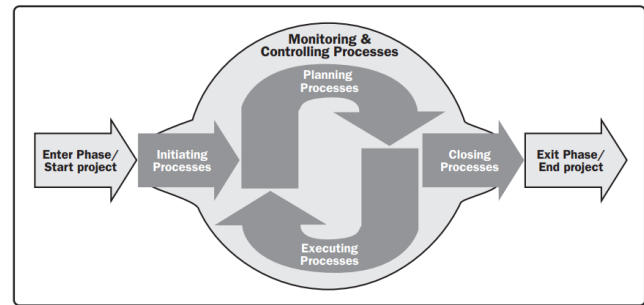


Fig. 1. Project management process groups [13].

2) The planning phase of a project involves establishing the project's scope, refining objectives, and determining the necessary actions to achieve those objectives [12]. This phase encompasses defining the course of action, making decisions, and formalizing the project's direction. Planning is a crucial step in project management, involving the conscious determination of actions to accomplish goals, and it is typically conducted after business planning and before project execution in fields such as construction [20].

3) The execution completing the tasks listed in the project management plan in order to comply with the project's criteria [12]. It encompasses the actual performance of project tasks and activities as outlined in the plan to fulfill the project's objectives [20].

4) The monitoring and controlling phase of project management involves tracking progress, reviewing performance, and making necessary changes to ensure project success [12]. Studies have emphasized the importance of timely control information and the ability to handle unexpected crises as critical factors in effective monitoring and control [20].

5) The closing phase of project management involves formally completing all activities and closing the project, phase, or contract [12]. Research conducted by Amponsah [21] focused on project failure/success factors in Ghana's agriculture, banking, and construction sectors. The study identified project management tools, techniques, and methods used by project managers and recommended that companies in Ghana focus on improving their project management activities for better outcomes.

### C. Challenges and Impact of AI Implementation in Project Management

Over time, project management has undergone significant changes and has become increasingly valuable to organizations. The collaboration between humans and machines in the face of major technological advancements presents significant opportunities for both businesses and individuals. One such advancement is Artificial Intelligence (AI), which is expected to greatly influence the role of project managers [6]. AI, despite its complexities, has the potential to

increase productivity and reduce errors in various fields, including software development projects. By providing insights into probable outcomes and removing extraneous information, AI improves project management and enables a focus on relevant facts [22].

The challenges of implementing AI in project management include barriers such as limited data availability, high costs of operation and equipment, and potential unemployment as AI replaces human workers. The successful adoption of AI requires technical staff with specialized skills and experience, along with a clear understanding of system integration and interoperability challenges [3]. While AI offers benefits like cost-effectiveness and reliability, it also raises issues of uncertainty and highlights both the positive and negative aspects of its adoption [23].

The adoption of AI in project management enhances decision-making quality by providing insights and support for potential outcomes. AI systems streamline information by eliminating redundancy, and auto-scheduling improves the robustness of project planning. Planning tools powered by AI, including hybrid computer systems, facilitate project control and objective configuration [8]. While contemporary management strategies like continuous alignment and agile approaches help overcome uncertainties, predictive analytics and machine learning contribute to better project outcomes in areas like KPIs, resource management, and estimate [3], [24].

### III. METHODOLOGY

This section explains the literature review process using the Systematic Literature Review (SLR) method. This method is used to evaluate and interpret all research related to the research questions, topic areas, or desired phenomena [25]. The goal of Systematic Literature Review is to generate accurate evaluations of the research topic using reliable, meticulous, and auditable methodologies. It is divided into three parts based on the Kitchenham [25] and references [26], [27]: Planning, Implementation, and reporting of SLR stages.

#### A. Planning the SLR

In this stage, keyword generation is conducted based on the main keywords and synonyms obtained at the beginning of the research, namely Artificial Intelligence and Project Management. Subsequently, these keywords are combined using "AND" to connect the main keywords and "OR" for each synonym or simpler word. The keywords and their synonyms or equivalents are as follows:

- “Artificial Intelligence”, synonym: “data mining”, “machine learning”
- “Project Management”

These keywords and their synonyms are combined to form the search keywords as follows:

(“artificial intelligence” AND “project management”) OR (“data mining” AND “project management”) OR (“machine learning” AND “project management”)

Then the search is conducted on several reputable literature and journal websites, based on title, abstract, or keywords. The

databases are IEEE Xplore, Springer Link, Emerald Insight, SAGE Journals, Science Direct, Scopus, and ACM Digital Library.

After determining which databases to use for the search, the researcher proceeds with the study selection process. This is intended to identify key studies within a research as direct evidence related to the previously obtained keywords [25]. The criteria for the study selection process in this research can be seen in Table I.

TABLE I. CRITERIA FOR STUDY SELECTION

Inclusion Criteria	IC1	Publications written in English
	IC2	Publications starting from 2018 until 2023
	IC3	Publications that truly focus on the keywords: artificial intelligence / data mining / machine learning and project management
	IC4	Publications that increasingly focus on the root problem, specifically discussing the challenges and impact of AI / DM / ML on project management
Exclusion Criteria	EC1	Research not written in English
	EC2	Research that discusses topics other than AI / DM / ML and project management
	EC3	Research published before 2018

#### B. Implementation of SLR

In the second phase of the Systematic Literature Review (SLR), the focus is on conducting an extensive search and selecting relevant literature. This involves identifying pertinent studies, extracting relevant information, and synthesizing the findings to obtain a comprehensive understanding of the research topic. Fig. 2 represents the flow diagram illustrating the selection process conducted from the initial data collection in various databases using the inclusion and exclusion criterias specified in Table I.

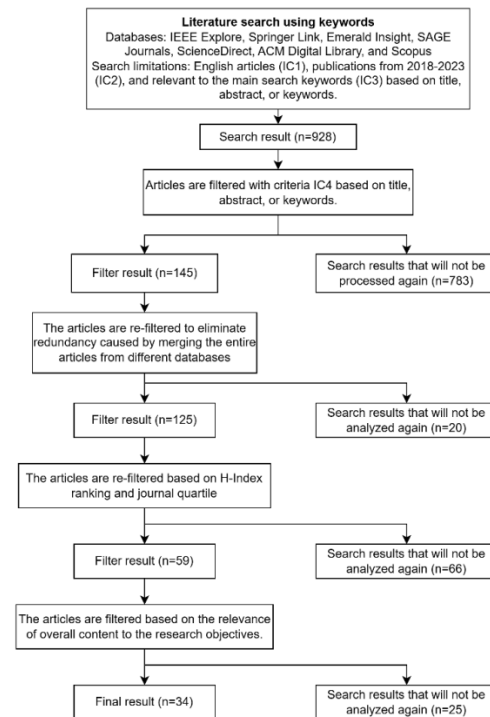


Fig. 2. Flow diagram of SLR implementation.

The total number of articles that will undergo further analysis related to the challenges and impacts on project management amounts to 34 articles, obtained from various high index journals and proceedings based on Scimago.

### C. Reporting the SLR

The final step of the Systematic Literature Review (SLR) is reporting the results. Table II presents a summary of all relevant article titles aligned with the research objectives, including information on the year and reference citation coding.

TABLE II. SUMMARY OF ALL RELEVANT ARTICLES

No	Title	Year	Index	Code
1	A BIM-data mining integrated digital twin framework for advanced project management	2021	Q1	[28]
2	A Machine Learning Study to Enhance Project Cost Forecasting	2022	Q3	[29]
3	A new hybrid ahp and dempster—shafer theory of evidence method for project risk assessment problem	2021	Q2	[30]
4	Activity classification using accelerometers and machine learning for complex construction worker activities	2021	Q1	[31]
5	An Approach Based on Bayesian Network for Improving Project Management Maturity: An Application to Reduce Cost Overrun Risks in Engineering Projects	2020	Q1	[32]
6	Application of Data Mining Technology in Field Verification of Project Cost	2021	Q4	[33]
7	Application of lean techniques, enterprise resource planning and artificial intelligence in construction project management	2019	Q4	[34]
8	Automated progress monitoring of construction projects using Machine learning and image processing approach	2022	Q2	[35]
9	Combined machine-learning and EDM to monitor and predict a complex project with a GERT-type network: A multi-point perspective	2023	Q1	[36]
10	Comprehensive project management framework using machine learning	2019	Q4	[37]
11	Data on Field Canals Improvement Projects for Cost Prediction Using Artificial Intelligence	2020	Q2	[38]
12	Data-driven project buffer sizing in critical chains	2022	Q1	[11]
13	Decision support system for final year project management	2019	Procd	[39]
14	Development and comparative of a new meta-ensemble machine learning model in predicting construction labor productivity	2022	Q1	[40]
15	DevOPs project management tools for sprint planning, estimation and execution maturity	2020	Q2	[41]
16	Empirically Exploring the Cause-Effect Relationships of AI Characteristics, Project Management Challenges, and Organizational Change	2021	Procd	[5]
17	Estimating production and warranty cost at the early stage of a new product development project	2021	Q3	[42]
18	Estimation of Risk Contingency Budget in Projects using Machine Learning	2022	Q3	[43]
19	Explainable machine learning for project management control	2023	Q1	[44]
20	Forecasting the scheduling issues in engineering project management: Applications of deep learning models	2021	Q1	[9]
21	Information Technology (IT) Governance Framework with Artificial Neural Network and Balance Scorecard to Improve the Success Rate of Software Projects	2022	Procd	[45]
22	Intelligent purchasing: How artificial intelligence can redefine the purchasing function	2021	Q1	[46]
23	Machine learning in project analytics: a data-driven framework and case study	2022	Q1	[47]
24	Project engineering management evaluation based on GABP neural network and artificial intelligence	2023	Q2	[48]
25	Project management: openings for disruption from AI and advanced analytics	2021	Q1	[49]
26	Proposal of a framework and integration of artificial intelligence to succeed IT project planning	2019	Q4	[50]
27	Recommendation of Project Management Practices: A Contribution to Hybrid Models	2022	Q1	[51]
28	Safety risk factors comprehensive analysis for construction project: Combined cascading effect and machine learning approach	2021	Q1	[52]
29	Symbiotic organisms search-optimized deep learning technique for mapping construction cash flow considering complexity of project	2020	Q1	[53]
30	The effectiveness of project management construction with data mining and blockchain consensus	2021	Q1	[54]
31	The impact of entrepreneurship orientation on project performance: A machine learning approach	2020	Q1	[55]
32	The value of data from construction project site meeting minutes in predicting project duration	2022	Q2	[56]
33	Using an Artificial Neural Network for Improving the Prediction of Project Duration	2022	Q2	[57]
34	Visual System Development for Construction Project Management by Using Machine Learning Algorithm	2022	Q2	[58]

Furthermore, the distribution of the scientific articles based on their publication years is presented in Fig. 3.

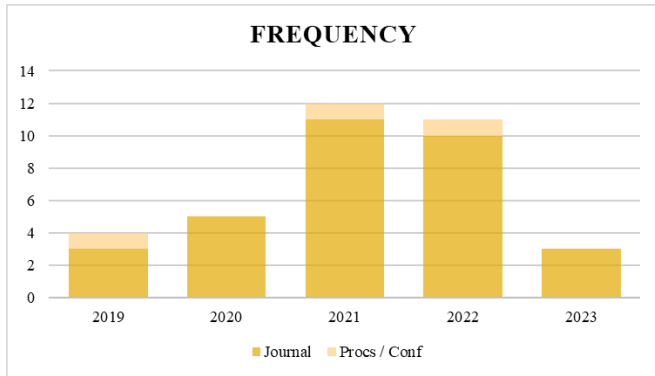


Fig. 3. Distribution of study per year.

#### IV. RESULTS AND ANALYSIS

In accordance with the objective of this research, a thorough mapping and analysis of the literature obtained will be conducted using the Systematic Literature Review (SLR) method described in the previous chapter. The outcomes obtained from this process will be divided into the mapping of AI tools used, challenges, and the impact of AI implementation on each article, as well as their summaries.

##### A. Mapping the AI tools, Implementation Challenges and Impacts based on Related Studies

Table III presents the further analysis results of the articles obtained in the previous SLR process, where the analysis is conducted regarding the AI tools used in the referenced studies, as well as the challenges and impacts of implementation in those studies. These AI tools can include algorithms, software, models, or others.

TABLE III. MAPPING THE AI TOOLS, CHALLENGES, AND IMPACT OF THE AI IMPLEMENTATION

Ref	AI Tools	Challenge	Impact
[28]	BIM, IoT, DM techniques, process mining, time series analysis, fuzzy miner algorithm, ARIMAX model	Real-time IoT data mapping and complexity management in large-scale projects.	AI enables early bottleneck detection, future workload prediction, and proactive decision-making for enhanced efficiency and adaptability.
[29]	Moving averages, schedule and cost performance factors	Nonlinear budget acquisition and cost growth patterns not accurately captured by traditional index-based models.	AI provides reliable and accurate cost estimates, improving project cost forecasting and maintaining positive stakeholder relationships.
[30]	Hybrid PCA-agglomerative unsupervised ML algorithm, Delphi method, statistical analysis, Taguchi Method	Uncertainty and variability of risk factors not accurately addressed by traditional decision-making methods.	AI quantifies risk factors for more informed project selection and improved decision-making, potentially enhancing project success rates.
[31]	Machine learning algorithms (e.g., K-Nearest Neighbors, Gradient Boosting), wearable accelerometers, feature selection techniques	Dealing with realistic scenarios, bias minimization, imbalanced datasets, and optimizing classifier performance with activity transitions.	AI enhances project decision-making, resource management, worker performance, and safety in construction, contributing to automation and process-level knowledge.
[32]	Bayesian networks	Integrating PMM models with Bayesian networks, formalizing expert knowledge, and addressing sector-specific factors and project drifts.	AI-based PMM diagnoses and predicts project performance failures, identifies drifts, and suggests corrective actions for comprehensive project success analysis.
[33]	Statistical data mining technology (not specified)	Developing a comprehensive organizational structure, implementing advanced management practices, improving cost communication, and addressing cost verification shortcomings.	AI introduces a data-driven approach, improving visualization, resource management, data collection, and cost control in project management.
[34]	Voice command, optical character recognition, LiDAR sensors	Integrating multiple systems, ensuring compatibility, overcoming resistance to change, and addressing complexity and uncertainty in construction projects.	AI minimizes waste, optimizes resource utilization, reduces costs, improves delivery time and quality, and enhances productivity in the construction industry.
[35]	Deep learning models, supervised CNN classifiers, image processing techniques	Reluctance, data standardization, and accuracy of automated tracking	Supervised CNN classifier improves activity recognition, enhancing supervision and progress tracking in construction projects.
[36]	Monte Carlo simulation, machine learning regression algorithms	Analyzing complex networks and selecting suitable algorithms	AI improves project monitoring, deviation identification, and completion time estimation, enhancing project performance and decision-making.
[37]	Artificial Intelligence (AI), machine learning models, online solutions (not specified)	Model selection, training, integration, data accuracy, and user interface development.	AI improves planning, decision-making, automation, and holistic solutions for project management.
[38]	Machine learning and AI techniques (Regression, Fuzzy, ANN, SVM, and others)	Dataset quality, algorithm selection, and interpretation/validation.	Understanding factors influencing project outcomes improves planning and resource allocation in sustainable projects.
[11]	Full-factor design of experiments, Monte Carlo simulation, regression techniques	Designing experiments, simulations, and ensuring model validity,	AI-based buffer sizing improves project planning and timeliness compared to traditional methods.
[39]	Naïve Bayes algorithm, JSP technology	Database organization, accuracy of Naïve Bayes, and technical issues.	AI platform for matching students with project topics and supervisors improves the quality and efficiency of final year research projects.
[40]	Ensemble machine learning algorithms	Preprocessing, algorithm selection, converting models, and case studies.	Accurate labor productivity prediction and influential factors guidance improve construction project management and overall productivity.



[41]	Real-time project data analysis, sprint estimation techniques, sentiment analysis	Real-time data analysis, accurate predictions, consensus, and sentiment analysis.	AI tools streamline project timelines, increase productivity, and improve decision-making in a DevOps environment.
[5]	AI techniques, machine learning algorithms, data analysis methods	Addressing AI's characteristics and implications in organizational settings.	Understanding the relationships between AI characteristics, project management practices, and organizational change improves AI implementation and decision-making.
[42]	Artificial neural network, linear regression techniques	Defining variables, acquiring data, and simulating cost variants.	Systematic consideration of cost factors and constraints improves cost estimation and optimization in new product development projects.
[43]	Machine learning algorithms, Monte Carlo simulation	Defining risks, selecting algorithms, and ensuring accuracy	Integration of machine learning techniques enhances Contingency Budget estimation accuracy and risk management in projects.
[44]	Monte Carlo simulation, statistical/machine learning models, explainable machine learning with SHAP	Integrating simulation and explainability, selecting models, and handling complexity.	Integration of Monte Carlo simulation, machine learning, and explainable machine learning improves project control and decision-making in uncertain environments.
[9]	LSTM and GRU models	Forecasting model selection, data quality, and interpretation.	LSTM and GRU models enable effective scheduling, resource allocation, and decision-making, improving overall project efficiency.
[45]	Artificial Neural Network (ANN) method, Balanced Score Card framework	Top management commitment, solution finding, and data availability.	AI integration in software project management enables informed decision-making and better outcomes, leading to the development of new project life-cycle solutions.
[46]	Automated systems for purchasing, and decision support	Resistance, integration, data quality, and training.	AI systems redefine purchasing roles, improve decision-making, supplier relations, and interdepartmental collaboration, highlighting the potential of AI capabilities in supplier management.
[47]	Machine learning algorithms (e.g., support vector machine, logistic regression, neural networks) using Python's Scikit-learn package	Limited evaluation, feature selection, and data imbalance.	AI-driven data analysis improves decision-making and understanding in construction project management.
[48]	BP neural network, genetic algorithms	Data availability, neural network limitations, and minimizing subjectivity.	AI-based system optimizes time, personnel, and resource utilization, improving engineering project management efficiency.
[49]	Software tools with AI and analytics features	Project managers adapting to software features	AI and analytics tools enhance project management with increased support, automation, and adaptive practices, emphasizing stakeholder relations and risk management.
[50]	Knowledge base, questioning system	Data acquisition, correctness, and risk management.	AI-based solution improves accuracy and efficiency of IT project planning, reducing failure rates and enabling cost savings.
[51]	Cluster analysis, association rule technique	Data collection, accuracy, and validation.	Guidance for selecting practices to enhance project management agility and effectiveness, suggesting avenues for further research in customizing hybrid models.
[52]	Machine learning, safety risk factor mining, AON networks	Abundant data, algorithm optimization, and quantifying risk impact.	AI improves safety risk analysis in construction projects, enhancing risk management, with further research needed for data independence and cost considerations.
[53]	Symbiotic organisms search (SOS) algorithm, LSTM, neural networks	Cash flow, model optimization, and data availability.	AI model accurately forecasts and controls cash flow in construction projects, improving cost management, decision-making, and resource allocation.
[54]	AI middle office, AI algorithms, blockchain technology, BIM	Data availability, integration, and consensus building.	Integration of AI middle office, BC technology, and BIM technology enhances trust, transparency, accuracy, security, and project management efficiency.
[55]	Machine learning algorithms (e.g., lasso, ridge, support vector machines, neural networks, random forest)	Algorithm selection, performance metrics, self-assessment, and generalizability.	AI techniques provide insights into operations and project management research, highlighting the potential of predictive analytics and entrepreneurial orientation for project success.
[56]	Data mining algorithms, random forest	Meeting minute data extraction, time-consuming capture, and data quality.	AI enables accurate project duration prediction, facilitating timely actions and improving project planning, leadership, and governance.
[57]	Artificial neural networks, genetic algorithms	Diverse datasets, adaptation to different organizations, and validation.	AI accurately predicts project duration for different organizations, adapting to various methods and datasets, enhancing project management decision-making.
[58]	Graph neural network, deep learning algorithms	Complex building data, accuracy of traditional algorithms, and limited early-stage.	Integration of complex data and machine learning algorithms improves building management efficiency, information access, communication, energy conservation, and safety.

Artificial intelligence tools in project management based on the related studies mentioned include BIM, IoT, DM techniques, process mining, time series analysis, fuzzy miner algorithm, ARIMAX model, moving averages, schedule and cost performance factors, hybrid PCA-agglomerative unsupervised ML algorithm, Delphi method, statistical analysis, Taguchi Method, machine learning algorithms (such as K-Nearest Neighbors and Gradient Boosting), wearable accelerometers, feature selection techniques, Bayesian

networks, voice command, optical character recognition, LiDAR sensors, deep learning models, supervised CNN classifiers, image processing techniques, Monte Carlo simulation, artificial neural networks, full-factor design of experiments, Naïve Bayes algorithm, ensemble machine learning algorithms, real-time project data analysis, sprint estimation techniques, sentiment analysis, software tools with AI and analytics features, knowledge base, questioning system, cluster analysis, association rule technique, safety risk factor mining, symbiotic organisms search (SOS) algorithm, LSTM,

AI middle office, lasso and ridge regression, support vector machines, random forest, data mining algorithms, genetic algorithms, and graph neural network.

From a project management standpoint, AI has the capability to replicate human cognitive functions such as decision-making and problem-solving [3]. These tools bring advanced capabilities to project management, enabling tasks such as data analysis, risk assessment, decision-making, performance monitoring, and optimization [42], [49]. They leverage AI, machine learning, and statistical techniques to improve project planning, resource allocation, cost estimation, risk mitigation, and overall project outcomes [43], [52]. These tools enhance efficiency, accuracy, and insights in project management processes, ultimately leading to improved project success rates and delivery [50], [54]

### B. Mapping AI Tools, Challenges, and Impacts of Articles based on PMBOK Process Group

Next is to map the process groups of PMBOK, which according to the literature are divided into 5, namely Initiating, Planning, Execution, Monitoring and Controlling, and Closing [12]. This mapping aims to integrate project management processes and product-oriented processes, ensuring proper coordination and alignment. The function of mapping the process group is to achieve this integration throughout the project lifecycle by tailoring the application of processes to meet the project's specific requirements and actively managing process interactions and tradeoffs to meet stakeholder needs [13]. The mapping results are presented in Table IV.

Based on the mapping results of the reference articles to the Process Groups in PMBOK, the next step involves reviewing the summaries of AI Tools used in each stage of the Process Group. Additionally, the challenges and their respective impacts on each process will be explored based on the conducted literature study.

TABLE IV. MAPPING RELATED STUDIES TO PROCESS GROUP

No	Process Group PMBOK	Reference	Freq
1	Initiating	[30], [38]	2
2	Planning	[5], [11], [29], [30], [32], [34], [37], [38], [41]–[43], [46]–[52], [54]	19
3	Execution	[34], [37], [39]–[41], [49]	6
4	Monitoring and Controlling	[9], [28], [31], [33]–[36], [44], [45], [49], [53]–[58]	16
5	Closing	[34], [49]	2

1) *Initiating process group*: In the Initiating process group of the PMBOK, both papers discussed the application of AI tools and analysis techniques to address challenges in risk management and cost prediction in projects.

The first paper [30] utilized a hybrid AHP and Dempster-Shafer methods along with various AI tools such as machine learning algorithms, Delphi method, statistical analysis, and Taguchi Method to tackle the challenge of uncertainty and variability in risk factors. These AI tools had a significant impact by offering a more reliable approach to project selection through the consideration and quantification of risk factors,

enabling informed decision-making and potentially improving project success rates.

Similar to the first article, the second paper [38] used Machine Learning and AI methods to estimate and forecast the cost and length of field canal development projects. These methods included regression, fuzzy logic, artificial neural networks, support vector machines, and others. Despite challenges related to dataset quality, algorithm selection, and interpretation/validation, the application of these techniques positively influenced the understanding of factors influencing project outcomes, leading to improved planning and resource allocation in sustainable projects.

In summary, within the Initiating process group, the integration of AI tools and analysis techniques in project management addresses challenges in risk assessment, cost prediction, and decision-making. These advancements provide more reliable approaches for project selection, considering and quantifying risk factors, and improving the overall success rates of projects. Additionally, the utilization of Machine Learning and AI techniques enhances the understanding of project outcomes and aids in better planning and resource allocation. However, challenges such as uncertainty, variability, and dataset quality need to be carefully addressed to ensure accurate predictions and interpretations in project management.

2) *Planning process group*: This comprehensive analysis summarizes the key findings and contributions of 19 scientific papers related to project management. These papers utilize various tools and techniques such as machine learning, artificial intelligence (AI), data mining, and statistical analysis to address challenges and enhance the process of group planning within the PMBOK framework. By leveraging these advanced tools, project managers can enhance decision-making, risk assessment, resource allocation, and project planning.

The application of these tools and methodologies in project management has significant impacts. By incorporating hybrid models, advanced algorithms, and statistical techniques, projects can benefit from improved risk assessment, more reliable project selection, accurate cost estimation, optimized resource utilization, and enhanced decision-making [29], [30], [38]. These advancements contribute to higher project success rates, improved project outcomes, and efficient resource allocation.

Furthermore, the integration of machine learning, artificial intelligence, and data analytics in project management processes enhances real-time data analysis, sentiment analysis, and decision support. This results in streamlined project timelines, increased productivity, and improved project management within a DevOps environment [41]. Additionally, the adoption of AI and advanced analytics tools in project management opens opportunities for disruption, automation, adaptive practices, and better stakeholder relations and risk management [32].

The construction industry also benefits from the application of AI, machine learning, and lean techniques. These

technologies enable the optimization of resource utilization, reduction of project costs, improvement of project delivery time and quality, and enhancement of overall productivity and efficiency [34], [52]. Moreover, the integration of AI in the South African construction project management industry contributes to job security, accident reduction, automation of high-risk tasks, and the creation of new job opportunities, thereby fostering regional development [59].

In summary, the 19 scientific papers examined showcase a wide range of tools and methodologies that address various challenges in project management. The application of these tools has a profound impact on risk assessment, project selection, cost estimation, resource allocation, decision-making, and overall project success rates. Additionally, the integration of AI and data analytics enhances real-time analysis, automation, adaptive practices, and stakeholder relations in project management processes. These advancements contribute to improved project outcomes, increased productivity, and regional development within the construction industry.

3) *Execution process group*: The analysis reveals valuable insights into the application of various tools, challenges, and impacts associated with the execution process group in PMBOK. These findings shed light on the utilization of tools such as lean techniques, enterprise resource planning (ERP), artificial intelligence (AI), machine learning models, and decision support systems to enhance project execution, productivity, process streamlining, and decision-making in different domains.

However, the implementation of these tools is not without challenges. The papers [34], [39] emphasize the challenges of integrating multiple systems, ensuring compatibility, overcoming resistance to change, addressing complexity and uncertainty, and dealing with technical issues. For example, a research [40], focuses on the challenges related to preprocessing, algorithm selection, converting models, and conducting case studies in the context of predicting construction labor productivity using ensemble machine learning models. These challenges necessitate comprehensive strategies and skillful implementation to successfully employ the identified tools in project execution.

Despite the challenges, the implementation of the identified tools brings about significant impacts on the execution process group. The papers highlight outcomes such as minimizing non-value-added efforts and waste, optimizing resource utilization, reducing project costs, improving project delivery time and quality, enhancing overall productivity and efficiency, and improving decision-making [34], [40], [41], [49]. For instance, a paper [34] discusses the impact of lean techniques, ERP, and AI in construction project management, which results in minimizing waste, optimizing resource utilization, reducing project costs, and improving project delivery time and quality within the construction industry.

Furthermore, the papers emphasize the potential of AI and advanced analytics tools in disrupting project management practices. These tools enhance productivity, support decision-making, automate processes, and emphasize stakeholder

relations and risk management [37], [49]. A paper specifically highlights the opening for disruption from AI and advanced analytics tools, underscoring their potential to transform project management practices through increased support, automation, and adaptive practices [49].

In summary, the analysis of the six papers highlights the significance of employing advanced tools and techniques in the execution process group in PMBOK. While the adoption of these tools offers numerous benefits, challenges related to system integration, resistance to change, and technical issues need to be addressed. The positive impacts resulting from their implementation include improved productivity, resource utilization, cost reduction, and enhanced decision-making. Further research and focused efforts are required to overcome the identified challenges and fully harness the potential of these tools for effective project execution.

4) *Monitoring and controlling process group*: The analysis of the 16 papers reveals significant insights into the utilization of tools, challenges, and impacts related to the monitoring and controlling process group in PMBOK. These papers highlight the application of various advanced tools and techniques such as Artificial Intelligence (AI), Building Information Modeling (BIM), Internet of Things (IoT), Data Mining (DM), machine learning, and image processing algorithms. These tools offer promising capabilities in enabling effective project control and monitoring in the construction domain. They facilitate complex data analysis, performance prediction, bottleneck detection, risk identification, and decision-making support.

However, the integration of these tools into project management practices is not without challenges. The papers emphasize the challenges associated with system integration, resistance to change, technology adoption, data accuracy, and project complexity [33], [34], [58]. The integration of multiple systems [45], [54], including voice command, optical character recognition, and LiDAR sensors, as mentioned in the certain paper [34], necessitates the overcoming of compatibility issues and resistance to change. Furthermore, the complexity and uncertainty inherent in construction projects pose additional challenges that demand comprehensive strategies and skilled implementation.

Despite these challenges, the implementation of the identified tools brings about significant impacts on the monitoring and controlling process group. The papers highlight outcomes such as enhanced efficiency, cost reduction, improved project quality, and timely delivery [9], [33], [34], [53], [57]. Moreover, the utilization of these tools fosters better decision-making, efficient resource management, improved stakeholder relations, and enhanced risk management [49], [57]. Other paper based on literature review [59] discusses the impact of AI integration in the construction industry, leading to improved job security, reduced accidents, automation of high-risk tasks, and the creation of new job opportunities.

In summary, the comprehensive analysis of the 16 papers underscores the significance of employing advanced tools and techniques for the monitoring and controlling process group in

PMBOK. While the adoption of these tools holds immense potential in project control and monitoring, challenges pertaining to system integration, resistance to change, technology adoption, and project complexity necessitate strategic interventions. However, the positive impacts resulting from their implementation include increased efficiency, cost reduction, improved quality, and informed decision-making, thereby driving effective project management. Further research and focused endeavors are required to address the identified challenges and fully harness the potential of these tools for improved project outcomes.

5) *Closing process group*: The two papers discussed the application of various tools and technologies in the Closing Process Group of project management according to the PMBOK. First paper [34] highlighted the use of tools like voice command, optical character recognition, and LiDAR sensors for what-if scenarios in construction project management. The challenges identified in this paper included integrating multiple systems, ensuring compatibility, overcoming resistance to change, and addressing complexity and uncertainty in construction projects. The impact of these tools was seen in minimizing non-value-added efforts, optimizing resource utilization, reducing project costs, improving project delivery time and quality, and enhancing overall productivity and efficiency within the construction industry.

Second paper [49] focused on the disruptions caused by AI and advanced analytics tools in project management. The tools discussed in this paper were software tools incorporating AI and analytics features. The main challenge highlighted was project managers adapting to these software features. However, the impact of these tools was significant, enhancing project management through increased support, automation, adaptive practices, and emphasis on stakeholder relations and risk management in the closing phase.

In summary, both papers emphasized the importance of utilizing advanced tools and technologies in the Closing Process Group of project management. The tools offer solutions to challenges such as system integration, resistance to change, and adaptation to software features. Moreover, the impacts mentioned, such as minimizing waste, optimizing resource utilization, reducing costs, improving quality, and enhancing productivity, align with the desired outcomes of the Closing Process Group in PMBOK.

## V. CONCLUSION

The findings of this systematic literature review (SLR) provide significant insights into the implementation of AI within the process groups defined in the PMBOK. The result highlights the importance of utilizing advanced tools and techniques in each process group and emphasizes the need to address specific challenges for successful AI implementation. Within the Initiating process group, AI tools and analysis techniques were effective in addressing challenges related to risk assessment, cost prediction, and decision-making. In the Planning process group, the application of AI and data analytics tools improved risk assessment, project selection, cost

estimation, resource allocation, and decision-making. The Execution process group experienced improved productivity, resource utilization, cost reduction, and enhanced decision-making through the use of advanced tools and techniques. The Monitoring and Controlling process group also demonstrated the potential of advanced tools and techniques in achieving increased efficiency, cost reduction, improved quality, and informed decision-making. Lastly, the Closing process group highlighted the value of incorporating AI, lean techniques, ERP, and machine learning, resulting in benefits such as cost reduction, efficiency improvement, and project closure success.

Overall, this SLR provides valuable insights for organizations aiming to integrate AI into their project management practices. By aligning the challenges and impacts with the PMBOK process groups, this research offers a structured framework that allows organizations to navigate the implementation of AI effectively. Understanding the specific challenges and impacts associated with each process group can help organizations address these issues and capitalize on the potential benefits of AI in project management, ultimately leading to improved project outcomes and increased productivity.

### A. Limitations of Study

The limitation of this study is that it should be noted that the authors have not explored or presented specific solutions to address the identified challenges discussed in the literature.

### B. Future Works

In future research, it is important to conduct a bibliographic analysis to identify relevant tools and map them into a specific framework. Additionally, creating a comprehensive mapping of the identified challenges within the framework will provide a better understanding of their origins and facilitate the search for effective solutions.

## REFERENCES

- [1] Taboada, A. Daneshpajouh, N. Toledo, and T. de Vass, "Artificial Intelligence Enabled Project Management: A Systematic Literature Review," *Applied Sciences (Switzerland)*, vol. 13, no. 8, 2023, doi: 10.3390/app13085014.
- [2] V. Maphosa and M. Maphosa, "Artificial Intelligence in Project Management Research: a Bibliometric Analysis," *J Theor Appl Inf Technol*, vol. 100, no. 16, pp. 5000–5012, 2022, doi: 10.5281/zenodo.7134073.
- [3] A. Alshaikhi and M. Khayyat, "An investigation into the Impact of Artificial Intelligence on the Future of Project Management," 2021 International Conference of Women in Data Science at Taif University (WiDSTaif ), 2021, doi: 10.1109/WiDSTaif52235.2021.9430234.
- [4] S. Bento, L. Pereira, R. Gonçalves, Á. Dias, and R. L. da Costa, "Artificial intelligence in project management: systematic literature review," *International Journal of Technology Intelligence and Planning*, vol. 13, no. 2, pp. 143–163, 2022, doi: 10.1504/ijtip.2022.126841.
- [5] C. Engel, P. Ebel, and B. van Giffen, "Empirically Exploring the Cause-Effect Relationships of AI Characteristics, Project Management Challenges, and Organizational Change," *Lecture Notes in Information Systems and Organisation*, vol. 47, no. February, pp. 166–181, 2021, doi: 10.1007/978-3-030-86797-3\_12.
- [6] C. Bodea, D. Rongui, O. Stanciu, and C. Mitea, *Artificial Intelligence impact in Project Management*. International project management association, 2020.

- [7] S. Yang, "A systematic literature review on the disruptions of artificial intelligence within the business world: in terms of the evolution of competences," no. June, pp. 0–39, 2022.
- [8] A. Belharet, U. Bharathan, B. Dzingina, N. Madhavan, C. Mathur, and Y.-D. B. Toti, "Report on the Impact of Artificial Intelligence on Project Management," pp. 1–53, 2020.
- [9] S. Liu and W. Hao, "Forecasting the scheduling issues in engineering project management: Applications of deep learning models," *Future Generation Computer Systems*, vol. 123, pp. 85–93, 2021, doi: 10.1016/j.future.2021.04.013.
- [10] A. T. T. Foster, "ARTIFICIAL INTELLIGENCE IN PROJECT MANAGEMENT.," *Cost engineering: a publication of the American Association of Cost Engineers.*, vol. 30, no. 6. The Association, Morgantown, WV ;, pp. 21–24. doi: info:doi/.
- [11] H. Li, Y. Cao, Q. Lin, and H. Zhu, "Data-driven project buffer sizing in critical chains," *Autom Constr*, vol. 135, no. December 2020, p. 104134, 2022, doi: 10.1016/j.autcon.2022.104134.
- [12] P. M. I. (ASV), *The standard for project management and a guide to the project management body of knowledge (PMBOK) 7th Edition.*, no. July. 2021.
- [13] PMI, *A Guide to the Project Management Body of Knowledge (PMBOK) Fifth Edition*, vol. 6. Project Management Institute, Inc., 2013.
- [14] B. Dickson, "What is artificial narrow intelligence (Narrow AI)?," 2020. <https://bdtechtalks.com/2020/04/09/what-is-narrow-artificial-intelligence-ani/> (accessed Jun. 12, 2023).
- [15] DAMA International, *DAMA-DMBOK: data management body of knowledge 2nd Edition*. 2017.
- [16] H. K. Mohamed, S. M. El-Debeiky, H. M. Mahmoud, and K. M. El Destawy, "Data mining for electrical load forecasting in Egyptian electrical network," in *2006 International Conference on Computer Engineering and Systems, ICCES'06*, 2006, pp. 460–465. doi: 10.1109/ICCES.2006.320491.
- [17] R. Shukla, P. Sharma, N. Samaiya, and M. Kherajani, "WEB USAGE MINING-A Study of Web data pattern detecting methodologies and its applications in Data Mining," 2020. doi: 10.1109/IDEA49133.2020.9170690.
- [18] S. Li, "Research on Data Mining Technology Based on Machine Learning Algorithm," *J Phys Conf Ser*, vol. 1168, no. 3, 2019, doi: 10.1088/1742-6596/1168/3/032132.
- [19] D. S. Hayes, "1999 International Student Paper Award Winner: Evaluation and Application of a Project Charter Template to Improve the Project Planning Process," *Project Management Journal*, vol. 31, no. 1, pp. 14–23, Mar. 2000, doi: 10.1177/875697280003100104.
- [20] A. P. Singh, "Project Management Process Group and Knowledge Area: Review," pp. 1–20, 2018, [Online]. Available: [www.ijarse.com](http://www.ijarse.com)
- [21] C. and P. M. ) Amponsah, Richard (School of Property, "Improving Project Management Practice in Ghana with Focus on Agriculture, Banking and Construction Sectors of the Ghanaian Economy," *Thesis*, vol. 8, no. July, p. 441, 2010, [Online]. Available: <https://researchbank.rmit.edu.au/view/rmit:10389>
- [22] N. O. C. Victor, "How Artificial Intelligence Influences Project Management," *Res Sq*, no. February, p. 19, 2023, doi: 10.21203/rs.3.rs-2535611/v1.
- [23] M. Chowdhury and A. W. Sadek, "Advantages and limitations of artificial intelligence," *Transportation Research Circular*, no. E-C168, pp. 6–8, 2012.
- [24] R. E. Levitt and J. C. Kunz, "Using artificial intelligence techniques to support project management," *AI EDAM*, vol. 1, no. 1, pp. 3–24, 1987, doi: DOI: 10.1017/S0890060400000111.
- [25] B. A. Kitchenham and S. Charters, "Guidelines for performing Systematic Literature Reviews in Software Engineering (Software Engineering Group, Department of Computer Science, Keele ...," *Technical Report EBSE 2007- 001*. Keele University and Durham University Joint Report, no. January, 2007.
- [26] T. Raharjo and B. Purwandari, "Agile project management challenges and mapping solutions: A systematic literature review," *ACM International Conference Proceeding Series*, pp. 123–129, 2020, doi: 10.1145/3378936.3378949.
- [27] 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, no. 2021, pp. 290–300, 2021, doi: 10.1016/j.procs.2021.12.143.
- [28] Y. Pan and L. Zhang, "A BIM-data mining integrated digital twin framework for advanced project management," *Autom Constr*, vol. 124, no. July 2020, p. 103564, 2021, doi: 10.1016/j.autcon.2021.103564.
- [29] T. Inan, T. Narbaev, and Ö. Hazir, "A Machine Learning Study to Enhance Project Cost Forecasting," *IFAC-PapersOnLine*, vol. 55, no. 10, pp. 3286–3291, 2022, doi: 10.1016/j.ifacol.2022.10.127.
- [30] S. M. Albogami, M. K. A. B. M. Ariffin, K. A. Ahmad, and E. E. B. Supeni, "A new hybrid ahp and dempster—shafer theory of evidence method for project risk assessment problem," *Mathematics*, vol. 9, no. 24, 2021, doi: 10.3390/math9243225.
- [31] L. Sanhudo et al., "Activity classification using accelerometers and machine learning for complex construction worker activities," *Journal of Building Engineering*, vol. 35, no. October 2020, 2021, doi: 10.1016/j.jobbe.2020.102001.
- [32] F. Sanchez, E. Bonjour, J.-P. Micaelli, and D. Monticolo, "An Approach Based on Bayesian Network for Improving Project Management Maturity: An Application to Reduce Cost Overrun Risks in Engineering Projects," *Comput Ind*, vol. 119, p. 103227, 2020, doi: <https://doi.org/10.1016/j.compind.2020.103227>.
- [33] Y. Jiang, "Application of Data Mining Technology in Field Verification of Project Cost," *Advances in Multimedia*, vol. 2021, 2021, doi: 10.1155/2021/3585878.
- [34] V. Vickranth, S. S. R. Bommarreddy, and V. Premalatha, "Application of lean techniques, enterprise resource planning and artificial intelligence in construction project management," *International Journal of Recent Technology and Engineering*, vol. 7, no. 6C2, pp. 147–153, 2019, [Online]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85067860685&partnerID=40&md5=b12e982df42fa911504460e636f9f6a2>
- [35] G. A.S. and J. B. Edayadiyil, "Automated progress monitoring of construction projects using Machine learning and image processing approach," *Mater Today Proc*, vol. 65, pp. 554–563, 2022, doi: 10.1016/j.matpr.2022.03.137.
- [36] A. Liang, L. Tao, and H. Lei, "Combined machine-learning and EDM to monitor and predict a complex project with a GERT-type network: A multi-point perspective," *Comput Ind Eng*, vol. 180, no. April, p. 109256, 2023, doi: 10.1016/j.cie.2023.109256.
- [37] K. S. N. Prasad and M. V. Vijaya Saradhi, "Comprehensive project management framework using machine learning," *International Journal of Recent Technology and Engineering*, vol. 8, no. 2 Special Issue 3, pp. 1373–1377, 2019, doi: 10.35940/ijrte.B1256.0782S319.
- [38] H. H. Elmousalami, "Data on Field Canals Improvement Projects for Cost Prediction Using Artificial Intelligence," *Data Brief*, vol. 31, p. 105688, 2020, doi: <https://doi.org/10.1016/j.dib.2020.105688>.
- [39] I. T. Afolabi, A. A. Adebisi, E. G. Chukwurah, and C. P. Igbokwe, "Decision support system for final year project management," in *Lecture Notes in Engineering and Computer Science*, 2019, pp. 233–237. [Online]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85083396939&partnerID=40&md5=dd33acf7d926df5a30ff59cfe97760ed>
- [40] I. Karatas and A. Budak, "Development and comparative of a new meta-ensemble machine learning model in predicting construction labor productivity," *Engineering, Construction and Architectural Management*, vol. ahead-of-p, no. ahead-of-print, Jan. 2022, doi: 10.1108/ECAM-08-2021-0692.
- [41] J. Angara, S. Prasad, and G. Sridevi, "DevOPs project management tools for sprint planning, estimation and execution maturity," *Cybernetics and Information Technologies*, vol. 20, no. 2, pp. 79–92, 2020, doi: 10.2478/cait-2020-0018.
- [42] M. Relich and I. Nielsen, "Estimating production and warranty cost at the early stage of a new product development project," *IFAC-PapersOnLine*, vol. 54, no. 1, pp. 1092–1097, 2021, doi: 10.1016/j.ifacol.2021.08.128.

- [43] C. Capone and T. Narbaev, "Estimation of Risk Contingency Budget in Projects using Machine Learning," *IFAC-PapersOnLine*, vol. 55, no. 10, pp. 3238–3243, 2022, doi: 10.1016/j.ifacol.2022.10.140.
- [44] J. I. Santos, M. Pereda, V. Ahedo, and J. M. Galán, "Explainable machine learning for project management control," *Comput Ind Eng*, vol. 180, no. April, 2023, doi: 10.1016/j.cie.2023.109261.
- [45] M. M. A. Ranesh, S. J. Samuel, R. Natchadalingam, and P. Jeyanthi, "Information Technology (IT) Governance Framework with Artificial Neural Network and Balance Scorecard to Improve the Success Rate of Software Projects," in *2022 6th International Conference on Electronics, Communication and Aerospace Technology*, 2022, pp. 1216–1221. doi: 10.1109/ICECA55336.2022.10009299.
- [46] O. Allal-Chérif, V. Simón-Moya, and A. C. C. Ballester, "Intelligent purchasing: How artificial intelligence can redefine the purchasing function," *J Bus Res*, vol. 124, no. October 2020, pp. 69–76, 2021, doi: 10.1016/j.jbusres.2020.11.050.
- [47] S. Uddin, S. Ong, and H. Lu, "Machine learning in project analytics: a data-driven framework and case study," *Sci Rep*, vol. 12, no. 1, 2022, doi: 10.1038/s41598-022-19728-x.
- [48] L. Yu, "Project engineering management evaluation based on GABP neural network and artificial intelligence," *Soft comput*, vol. 27, no. 10, pp. 6877–6889, 2023, doi: 10.1007/s00500-023-08133-9.
- [49] F. Niederman, "Project management: openings for disruption from AI and advanced analytics," *Information Technology & People*, vol. 34, no. 6, pp. 1570–1599, Jan. 2021, doi: 10.1108/ITP-09-2020-0639.
- [50] R. Hassani and Y. El Bouzekri El Idrissi, "Proposal of a framework and integration of artificial intelligence to succeed IT project planning," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 8, no. 6, pp. 3396–3404, 2019, doi: 10.30534/ijatcse/2019/114862019.
- [51] M. J. Bianchi, E. C. Conforto, E. Rebentisch, D. C. Amaral, S. O. Rezende, and R. de Pádua, "Recommendation of Project Management Practices: A Contribution to Hybrid Models," *IEEE Trans Eng Manag*, vol. 69, no. 6, pp. 3558–3571, 2022, doi: 10.1109/TEM.2021.3101179.
- [52] G. Ma, Z. Wu, J. Jia, and S. Shang, "Safety risk factors comprehensive analysis for construction project: Combined cascading effect and machine learning approach," *Saf Sci*, vol. 143, p. 105410, 2021, doi: <https://doi.org/10.1016/j.ssci.2021.105410>.
- [53] M. Y. Cheng, M. T. Cao, and J. G. Herianto, "Symbiotic organisms search-optimized deep learning technique for mapping construction cash flow considering complexity of project," *Chaos Solitons Fractals*, vol. 138, 2020, doi: 10.1016/j.chaos.2020.109869.
- [54] W. Li, P. Duan, and J. Su, "The effectiveness of project management construction with data mining and blockchain consensus," *J Ambient Intell Humaniz Comput*, 2021, doi: 10.1007/s12652-020-02668-7.
- [55] S. Sabahi and M. M. Parast, "The impact of entrepreneurship orientation on project performance: A machine learning approach," *Int J Prod Econ*, vol. 226, no. April 2019, p. 107621, 2020, doi: 10.1016/j.ijpe.2020.107621.
- [56] J. van Niekerk, J. Wium, and N. de Koker, "The value of data from construction project site meeting minutes in predicting project duration," *Built Environment Project and Asset Management*, vol. 12, no. 5, pp. 738–753, 2022, doi: 10.1108/BEPAM-03-2021-0047.
- [57] I. Lishner and A. Shtub, "Using an Artificial Neural Network for Improving the Prediction of Project Duration," *Mathematics*, vol. 10, no. 22, 2022, doi: 10.3390/math10224189.
- [58] X. Huang and M. Liang, "Visual System Development for Construction Project Management by Using Machine Learning Algorithm," *Optik (Stuttg)*, p. 170460, 2022, doi: 10.1016/j.ijleo.2022.170460.
- [59] S. Makaula, M. Munsamy, and A. Telukdarie, "Impact of artificial intelligence in South African construction project management industry," in *Proceedings of the International Conference on Industrial Engineering and Operations Management*, 2021, pp. 148–162. [Online]. Available: <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85121110767&partnerID=40&md5=317512e166c09d24a3f00e4eefa0645a>

# Deep Residual Convolutional Long Short-term Memory Network for Option Price Prediction Problem

Artur Dossatayev<sup>1</sup>, Ainur Manapova<sup>2</sup>, Batyrkhan Omarov<sup>3</sup>  
International Information Technology University, Almaty, Kazakhstan<sup>1,3</sup>  
Academy of Civil Aviation, Almaty, Kazakhstan<sup>2</sup>  
NARXOZ University, Almaty, Kazakhstan<sup>3</sup>  
Al-Farabi Kazakh National University, Almaty, Kazakhstan<sup>3</sup>

**Abstract**—In the realm of financial markets, the precise prediction of option prices remains a cornerstone for effective portfolio management, risk mitigation, and ensuring overall market equilibrium. Traditional models, notably the Black-Scholes, often encounter challenges in comprehensively integrating the multifaceted interplay of contemporary market variables. Addressing this lacuna, this study elucidates the capabilities of a novel Deep Residual Convolution Long Short-term Memory (DR-CLSTM) network, meticulously designed to amalgamate the superior feature extraction prowess of Convolutional Neural Networks (CNNs) with the unparalleled temporal sequence discernment of Long Short-term Memory (LSTM) networks, further augmented by deep residual connections. Rigorous evaluations conducted on an expansive dataset, representative of diverse market conditions, showcased the DR-CLSTM's consistent supremacy in prediction accuracy and computational efficacy over both its traditional and deep learning contemporaries. Crucially, the integration of residual pathways accelerated training convergence rates and provided a formidable defense against the often detrimental vanishing gradient phenomenon. Consequently, this research positions the DR-CLSTM network as a pioneering and formidable contender in the arena of option price forecasting, offering substantive implications for quantitative finance scholars and practitioners alike, and hinting at its potential versatility for broader financial instrument applications and varied market scenarios.

**Keywords**—Deep learning; CNN; LSTM; prediction; option price

## I. INTRODUCTION

Option pricing, an intrinsic component of financial markets, serves as the fulcrum upon which significant economic decisions, from individual investments to institutional strategies, hinge [1]. These derivative contracts, which bestow upon the holder the right, but not the obligation, to buy or sell an underlying asset at a specified price before a predetermined date, play a pivotal role in portfolio diversification, risk hedging, and speculative ventures [2]. Historically, the Black-Scholes model has been emblematic in the realm of option pricing, a seminal formula that established a theoretical framework for determining the fair market value of a European-style option [3]. However, as financial markets evolved, becoming increasingly complex and intertwined, driven by a plethora of factors ranging from geopolitical

dynamics to technological innovations, the limitations of such traditional models have become conspicuously palpable.

The metamorphosis of financial markets into data-dense ecosystems, replete with myriad variables and indicators, necessitates predictive models with a capacity for high-dimensional data processing and the discernment of intricate temporal relationships [4]. The recent resurgence in artificial intelligence, more specifically in deep learning, offers a promising vista for addressing these challenges. Deep learning architectures, distinguished by their hierarchical structure and ability to autonomously extract salient features from raw data, have been progressively permeating various domains, ranging from image recognition to natural language processing [5]. Within the financial sphere, their application promises to harness the vast expanses of data to render more nuanced and accurate predictions.

In this milieu, the Deep Residual Convolution Long Short-term Memory (DR-CLSTM) network emerges as a potent amalgamation of several cutting-edge deep learning paradigms. By intertwining the feature extraction capabilities of Convolutional Neural Networks (CNNs) [6] with the temporal relationship discernment offered by Long Short-term Memory (LSTM) networks [7], and further augmenting this synergy with deep residual connections, the DR-CLSTM aspires to provide a holistic solution to the option price prediction quandary. The convolutional layers, renowned for their prowess in spatial hierarchies discernment, sieve through vast datasets, isolating pertinent features integral to option pricing. Concurrently, the LSTM layers, celebrated for their ability to capture long-term dependencies by combating the vanishing gradient problem, harness these features to forecast temporal sequences, thereby rendering predictions [8]. The addition of deep residual connections further augments this architecture. By allowing activations to bypass one or more layers, these connections expedite the training process, ensuring faster convergence and fortifying the network against potential gradient diminution.

This paper seeks to explore and substantiate the efficacy of the DR-CLSTM network in the realm of option price prediction. In doing so, it aims to contribute a novel tool to the arsenal of quantitative finance, fostering enhanced market efficiency, informed investment decisions, and a deeper

comprehension of the intricate tapestry of factors influencing option prices. Furthermore, by juxtaposing the DR-CLSTM with both traditional computational models and contemporary deep learning architectures, this study endeavors to provide a holistic perspective on the trajectory of option price prediction methodologies, underlining the transformative potential of hybrid deep learning structures.

The ensuing sections will delve into the theoretical underpinnings of the DR-CLSTM, elucidating the individual components and their synergistic interplay. This will be followed by a comprehensive methodology section, detailing the dataset employed, the evaluation metrics, and the comparative models. Subsequent sections will present the empirical findings, discussions, and potential implications for both academia and industry. The paper will culminate with a conclusion, encapsulating the key insights gleaned and charting potential avenues for future research in this dynamic and ever-evolving domain.

## II. RELATED WORKS

The ever-evolving landscape of financial modeling and forecasting has witnessed a plethora of innovations and methodologies over the years. As we delve into the intricate world of option price prediction, it is imperative to ground our discussions within the context of prior research and explorations in this domain [9]. The "Related Works" section endeavors to provide readers with a comprehensive overview of seminal works, pioneering methodologies, and noteworthy contributions that have shaped the trajectory of this field. By juxtaposing our current study with these foundational works, we aim to highlight both the advancements made and the gaps that our research seeks to bridge. Let us embark on this journey of retrospection, understanding the milestones that have been achieved and setting the stage for the novel contributions of our study.

### A. Traditional Models for Option Pricing

Option pricing, a central facet of financial mathematics, has been subject to rigorous academic scrutiny for decades. It revolutionized this arena with their eponymous model, providing a closed-form solution for European option pricing [10]. The Black-Scholes model, predicated on certain assumptions such as constant volatility and interest rates, quickly became a cornerstone of financial markets. However, the assumptions underlying the model often diverge from real-world market conditions, leading to potential mispricing [11]. This inherent limitation paved the way for alternative stochastic volatility models, which attempt to address the constant volatility constraint [12].

### B. Emergence of Machine Learning in Financial Forecasting

With the exponential growth of computational capabilities and the proliferation of vast financial datasets, machine learning has transitioned from a theoretical concept to an instrumental tool in financial forecasting. Over the past two decades, the field has seen a marked departure from traditional econometric models towards more adaptive and self-learning algorithms. Next research advanced the argument that neural networks possess the capacity to model complex non-linear relationships inherent in financial markets, a dimension often

inadequately captured by traditional methods [13]. Next study specifically applied Recurrent Neural Networks (RNNs) to the domain of option pricing, highlighting the technology's aptitude for capturing intricate temporal sequences [14]. This evolution indicates a paradigmatic shift in financial forecasting methodologies, illustrating the potential of machine learning techniques to address the increasingly multifaceted nature of financial markets.

### C. Deep Learning and Financial Markets

In the expansive sphere of machine learning, deep learning stands out, characterized by its multi-layered neural networks adept at handling high-dimensional data. Originating from image and video recognition tasks, as underscored by LeCun et al. (2015), these algorithms have witnessed a significant adaptation to financial analytics [15]. The strength of Convolutional Neural Networks (CNNs) lies in their autonomous feature extraction capabilities, which have found resonance in financial time series analysis. A new study successfully applied CNNs to forecast stock price movements, underscoring their superiority over conventional methods [16]. This adaptation of deep learning to finance not only signifies its versatility but also heralds a new era in financial forecasting. Embracing these sophisticated architectures promises a more nuanced understanding of financial market intricacies, catalyzing more informed and data-driven decision-making processes in the sector.

### D. Residual Networks in Deep Learning

Within the intricate tapestry of deep learning architectures, Residual Networks (ResNets) have carved a distinctive niche. The transformative nature of ResNets lies in their "shortcut connections", which allow activations to bypass certain layers, facilitating the learning of identity functions [17]. This innovative approach, originally tailored for visual tasks, has significantly mitigated challenges associated with training deeper neural networks, predominantly by averting performance degradation. Its adaptation to financial applications, though still burgeoning, shows immense potential. The principal virtue of these networks is their ability to combat the notorious vanishing gradient problem, thus enhancing the depth and complexity of models without sacrificing accuracy. The foray of ResNets into financial forecasting stands as testament to the evolving landscape of deep learning, offering fresh perspectives and tools to navigate the multifarious nature of financial data.

### E. Hybrid Deep Learning Architectures

As the deep learning domain continues its relentless evolution, the emergence of hybrid architectures signifies a pivotal juncture. These models, synergizing distinct deep learning techniques, aim to harness the individual strengths of each component, resulting in a more comprehensive and potent analytical tool. Gosztolya et al. (2017) pioneered in this space, presenting a confluence of Convolutional Neural Networks (CNNs) and Long Short-term Memory networks (LSTMs) for nuanced time series forecasting [18]. Their work highlighted the hybrid model's capability to simultaneously capture spatial features and temporal dynamics. Yet, the realm of hybrid architectures is expansive, with the potential for myriad combinations. As these integrated frameworks gain traction,



they promise to redefine the boundaries of what deep learning can achieve, paving the way for more sophisticated, adaptable, and accurate solutions in diverse application areas, including finance.

#### F. Challenges in Option Price Prediction

Option price prediction, central to financial analysis, is fraught with complexities. Despite the sophistication of existing models, accurate forecasting remains an elusive goal, owing to the non-stationary nature of financial markets. A multitude of variables, both foreseen and unforeseen, continually impact option prices, making their behavior highly unpredictable. Traditional models, exemplified by Black-Scholes, while groundbreaking, have faced criticism for their stringent assumptions, as highlighted [19]. Emerging machine learning models, despite their adaptability, are not immune to challenges, notably overfitting and susceptibility to market anomalies. Dhiman et al., (2020) elucidate these limitations, stressing the need for models that are both adaptive and robust [20]. As the financial landscape grows increasingly intricate, the pursuit of a holistic, accurate, and resilient option pricing model remains a quintessential challenge, beckoning further research and innovation.

#### G. Gap in Current Literature

Within the vast expanse of academic literature dedicated to financial forecasting and particularly option price prediction, certain gaps persistently emerge. The trajectory of research has indeed traversed from traditional mathematical formulations to sophisticated computational architectures, yet complete solutions appear to remain on the horizon. While many studies, such as those by Black and Scholes, have laid foundational pillars, and others have delved deep into the capabilities of machine learning and deep learning techniques, a comprehensive amalgamation seems somewhat elusive.

Most evident is the lack of extensive exploration into hybrid deep learning frameworks, particularly in their application to financial markets. The combination of multiple neural architectures, each with its inherent strengths, presents a promising frontier. Kumar et al. (2023) offered a glimpse into the potential of these combined structures, but the literature remains scant in its entirety [21].

Furthermore, while there is an abundance of studies leveraging individual deep learning structures, the integration of residual connections within these architectures is a relatively uncharted territory, especially in financial forecasting contexts [22]. Current methodologies exhibit a propensity to lean heavily towards either feature extraction or temporal sequence modeling. Rare are the models that harmoniously intertwine both facets.

This paucity in holistic models underscores the pivotal gap in the existing literature. The quest remains for a unified model, like the DR-CLSTM, that seamlessly melds the strengths of various deep learning techniques, thus addressing the multifaceted challenges of option price prediction. This present study aims to contribute to bridging this gap, offering a fresh perspective grounded in both retrospective analyses and forward-looking innovations.

The trajectory of option price prediction methodologies has witnessed a paradigm shift from traditional mathematical models to sophisticated computational frameworks, epitomized by deep learning techniques [23]. As financial markets continue to evolve, becoming increasingly intricate, the need for robust, adaptive, and comprehensive models becomes paramount. The DR-CLSTM network, as explored in this paper, emerges in response to this exigency, grounded in a rich tapestry of academic research spanning diverse domains. By weaving together the strengths of CNNs, LSTMs, and ResNets, this study aims to contribute a novel perspective to the discourse on option price prediction, anchored in both historical precedents and contemporary innovations.

### III. MATERIALS AND METHODS

The crux of any research endeavor lies in the robustness of its methodologies and the quality of the materials employed. In this "Materials and Methods" section, we elucidate the systematic approaches, tools, and datasets harnessed to facilitate our investigative journey into option price prediction. Serving as the backbone of our study, this section ensures replicability and offers a transparent window into the foundational processes that underpin our findings. Herein, we meticulously detail the data sources, the preprocessing techniques adopted, and the intricacies of the analytical methods applied. By offering this in-depth exposition, we aim to provide a roadmap for researchers, practitioners, and enthusiasts alike, enabling a clear understanding of the mechanisms that drive our research forward. Let us navigate through the critical phases and elements that constitute the scientific rigor of our study.

#### A. Problem Statement

In the realm of derivative pricing, a model is recognized as being aligned with a set of benchmarking instruments when the calculated parameters, within its structure, coincide with prevailing market values. Calibration can be conceptualized as the act of juxtaposing a model against these benchmarking entities. The act of defining parameters to affirm measuring conditions encapsulates the essence of model calibration [24]. While derivative pricing may be perceived as the prospective challenge, calibration can be seen as its antithesis. Hypothetically, given the prices of Call options across all strikes and durations, the calibration quandary can be directly tackled through an inverse formulation. Yet, confronted with a restricted set of derivative valuations, calibration becomes an indeterminate challenge, necessitating the deployment of regularization techniques in real-world settings.

The integration of machine learning in predicting option prices has demonstrated potential, enhancing model precision, agility, and resilience. Notwithstanding, existing models grapple with obstacles such as the requisite for vast and varied datasets, heightened sensitivity to inputs and settings, and the intricacy of encapsulating intricate interactions affecting the option price [25]. Subsequent inquiries should prioritize crafting resilient and comprehensible models, assimilating supplementary data avenues, external variables, and broadening the application horizons of machine learning within financial domains.

Conceptually, this mirrors determining the apex solution for a given conundrum. Supposing we possess a roster of Call options market valuations  $C_M$  for specific durations  $T_i$ ,  $i = 1, \dots, n$  and strike values  $K_{i,j}$ ,  $j = 1, \dots, m_i$  which align with prevailing market metrics like  $S$ ,  $r$ ,  $q$ , among others. Given this classification, the assortment of exchanged strikes  $K_{i,j}$  may differ based on the duration  $T_i$ . If we take into account calibrating the model  $M$  delineated in Section I, it encompasses model parameters  $p$  in conjunction with input data metrics. We differentiate  $p$  and based on their intrinsic implications: while parameters  $p$  is derived through calibration, the input metrics symbolize discernible market constants, such as  $S$ ,  $r$ ,  $q$ ,  $T$ , and  $K_3$ .

Given the relevant inputs for  $p$  and, the model  $M$  can formulate the mathematical framework for Call option valuations  $C$ . To ascertain the model's calibrated parameters, it's imperative to resolve the associated optimization conundrum:

$$\arg \min_p \sum_{i=1}^n \sum_{j=1}^{m_i} w_{i,j} \| C_M(\theta_{i,j}) - C(\theta_{i,j}, p) \quad (1)$$

In this context,  $i$  and  $j$  represent weight coefficients, and  $L_2$  stands as a normative measure for  $k$  and  $k$ , employed in the formulation given by Eq. (1). Broadly, to address the intricacies of the non-linear function, it becomes imperative to adopt global optimization strategies. Such a need arises as the multi-faceted objective (or loss) function could potentially harbor multiple localized minima. Moreover, this challenge may be inherently bounded, a phenomenon frequently observed when deriving the implied volatility surface from market data points.

### B. The Proposed Model

Financial assets intrinsically exhibit stochastic behavior, necessitating representation through intricate nonlinear and multivariate functions, thus rendering option pricing an exceptionally complex challenge. Traditional approaches often rely heavily on numerous statistical presuppositions concerning market or data dynamics. In stark contrast, deep learning excels in its capacity to model nonlinearities, accomplishing this without such restrictive assumptions. This section succinctly outlines the non-parametric models employed in the referenced studies.

In the initial publication of this research, a Convolutional-LSTM model was introduced to address the dual complexities of spatial and temporal data modeling. Fig. 1 shows the proposed Convolutional LSTM network for option price prediction. The data architecture remains consistent here, interpretable as encompassing  $C$  channels, given its bidimensional configuration. Thus, the comprehensive dataset serving as input adheres to the structure [26]. The 2D convolution operation is executed across the  $D$  and  $E$  parameters, spanning each row and layers of the representation.

Alternatively, the dataset related to the option lacks a dimension, crucial for effective Convolutional-LSTM processing. Its sole permissible structure is  $(T;C;N;D)$ . A pragmatic resolution involves adapting the Convolutional-LSTM algorithm to harness 1D convolution, foregoing the 2D pooling layers. This unidimensional convolutional operation predominantly traverses the  $D$  component. Such convolution is consistently applied across pertinent inputs, cell outputs, hidden outputs, and gates. When centered on discerning correlations within market data, this 1D convolution proves markedly advantageous, bolstering performance over mere vector-based input methodologies.

The computational process unfolds in the subsequent manner delineated herein:

$$\begin{aligned} i_t &= \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} * C_{t-1} + b_i) \\ f_t &= \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} * C_{t-1} + b_f) \\ C_t &= f_t \circ C_{t-1} + i_t \circ \tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_0) \\ H_t &= o_t \circ \tanh(C_t) \end{aligned} \quad (2)$$

Where the symbol  $*$  denotes the unidimensional convolution occurring between the latent phase and the convolutional kernels.

For the purpose of ensuring coherence in comparison, the input data is further reshaped to adopt the following configuration:  $(N; 1; T; C D) = (N; 1; 10; 15)$ . In essence, this consolidates all parameters into a unified channel. This modification aims to test our methodology's efficacy in feature representation

In our research, we utilized data sourced from the Chinese stock and fund markets between April 2018 and June 2020, with a specific emphasis on 50ETF option and stock option data [27]. Initially, the dataset comprises 55,047 entries and 34 distinct variables. We focused on selections with an effective duration of a minimum of 20 days, resulting in 829 such choices. Of these, the first 600 are designated for training, with the subsequent 229 allocated for testing purposes. This data is reshaped according to the specifications detailed in Section 3(B), resulting in a final training set with 33,210 entries and a test set with 11,386 entries. The data for this study was procured from Yahoo Finance [28].

Prior to the commencement of our experiments, it was imperative to normalize each variable within the dataset. This normalization is essential to counteract potential imbalances or asymmetries in the neural network. For illustration:

$$\begin{aligned} \tilde{x} &= (x - \mu) / \sigma \\ \mu &= \frac{1}{n} \sum_{i=1}^N x_i, \quad \sigma^2 = \frac{1}{n} \sum_{i=1}^N (x_i - \mu)^2 \end{aligned} \quad (3)$$

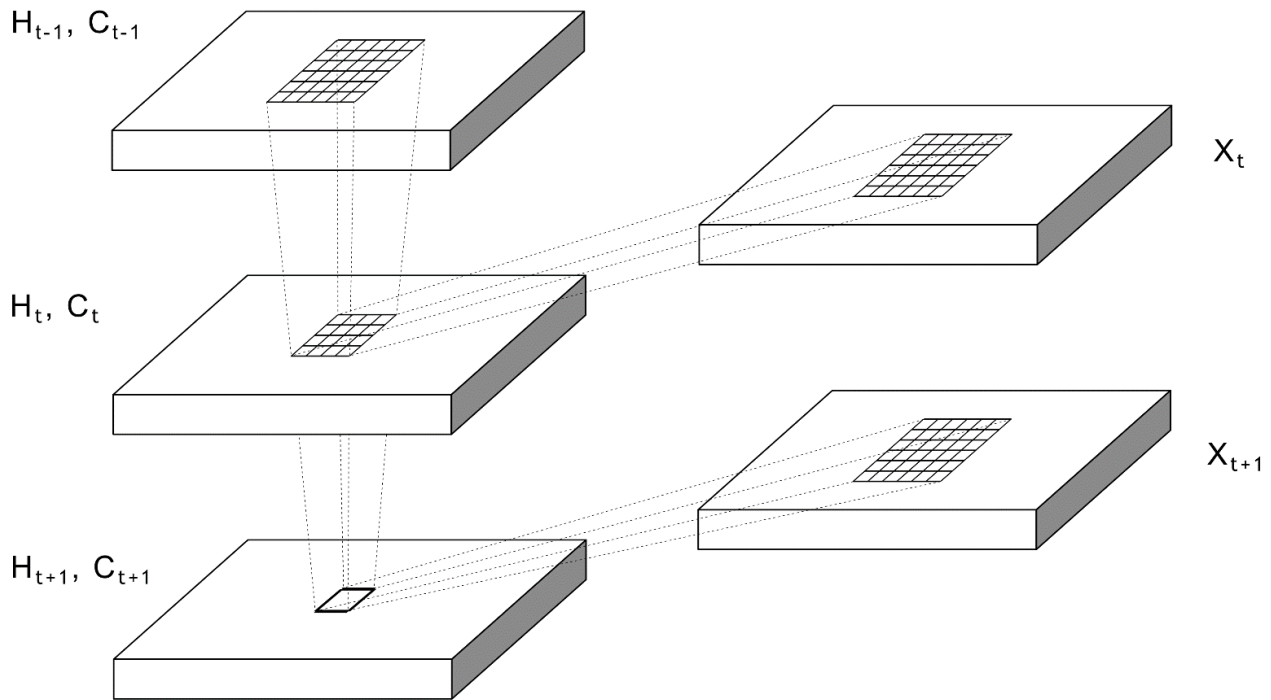


Fig. 1. The proposed convolutional-LSTM network for option price prediction.

#### IV. EXPERIMENTAL RESULTS

In this section, we present the empirical outcomes of our study, offering a rigorous evaluation of the conducted experiments. The results elucidated herein provide a comprehensive insight into the efficacy of the proposed models in light of the research objectives. By scrutinizing these outcomes, readers can glean an understanding of the model's robustness, accuracy, and adaptability in various scenarios. Moreover, the results will be juxtaposed with established benchmarks and previous studies, serving as a comparative framework. This comparative analysis aims to underscore the advancements and potential shortcomings of the current research. Without further ado, let us delve into the detailed exposition of the experimental findings.

##### A. Evaluation Metrics

Mean Squared Error (MSE) stands as one of the paramount metrics in quantitative assessment, especially when the objective is to measure the average magnitude of error between predicted and actual observations [29]. Mathematically defined as the average of the squared differences between the forecasted and observed values, MSE serves as an indicator of the accuracy of a model's predictions. One of its distinctive characteristics is its amplification of larger errors, due to the inherent squaring of discrepancies. Consequently, models with a lower MSE are preferred as they suggest a closer fit to the actual data. However, it's essential to note that while MSE is profoundly informative, its sensitivity to outliers can sometimes overemphasize large errors. As such, it is often considered in conjunction with other evaluation metrics to provide a more comprehensive assessment of a model's performance.

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (4)$$

Mean Absolute Error (MAE) is a critical metric employed in the realm of predictive modeling to gauge the average magnitude of errors between forecasted and actual observations [30]. Unlike the Mean Squared Error (MSE), which squares discrepancies, the MAE takes the absolute value of these errors. As a result, it provides a linear score where all individual differences have equal weight. This ensures that the metric is not disproportionately influenced by outliers, rendering it less sensitive to large deviations compared to MSE. Essentially, the MAE quantifies the average vertical distance between each point and the identity line in a prediction plot. Lower MAE values indicate a model that is adept at making predictions closely aligned with actual outcomes. Given its intuitive nature and resistance to the undue influence of outliers, MAE often serves as a pivotal parameter in many evaluation frameworks, particularly when the objective is to obtain a straightforward understanding of prediction accuracy.

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (5)$$

Root Mean Squared Error (RMSE) stands as a prominent metric in predictive analytics and modeling, serving to evaluate the magnitude of error between predicted and actual outcomes [31]. Derived from the Mean Squared Error (MSE), the RMSE is computed by taking the square root of the averaged squared differences between forecasted and observed values. This operation preserves the unit of the measurements, facilitating a more intuitive interpretation of the error magnitude. By emphasizing larger errors due to its squared components,

RMSE is particularly sensitive to significant discrepancies and outliers. As such, a lower RMSE denotes a model's superior predictive accuracy, suggesting its predictions are in closer proximity to observed data. However, given its sensitivity, RMSE is often juxtaposed with other metrics, such as the Mean Absolute Error (MAE), to ensure a well-rounded assessment of a model's performance, especially in datasets with pronounced outliers.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (6)$$

Mean Absolute Percentage Error (MAPE) is a quintessential metric in the analytical domain, primarily utilized to gauge the accuracy of forecasting methods in terms of percentage [32]. It calculates the average absolute percentage discrepancy between observed and predicted values relative to the actual value. This metric offers a scale-independent perspective on errors, enabling comparisons across varied units and magnitudes. A salient feature of MAPE is its intuitive interpretation, as it directly quantifies the prediction error as a percentage, facilitating easy comprehension of the model's performance in practical terms. Lower MAPE values are indicative of superior predictive accuracy, suggesting that predictions closely mirror actual observations. However, one caveat associated with MAPE is its potential to produce undefined or infinite values when the actual observation is zero. Moreover, it can sometimes disproportionately penalize underestimations compared to overestimations. Despite these nuances, MAPE remains a favored choice in many scenarios, especially when

stakeholders seek a percentage-based evaluation of predictive accuracy.

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (7)$$

Beyond the aforementioned metrics, it's imperative to conduct comprehensive model assessments employing methods like cross-validation, testing on unseen samples, and juxtaposing the model's outputs against conventional approaches or established benchmarks. Such rigorous evaluations bolster the model's resilience and its capacity to generalize across diverse datasets, thereby enhancing the precision and dependability of option price forecasts in real-world financial scenarios.

### B. Experimental Results

In this section, we delve into the analytical outcomes of the suggested model as elaborated in the preceding segment. Fig. 2 delineate the efficacy of the ConvLSTM model in forecasting option prices, juxtaposing the accuracy during training and validation phases across 200 epochs. The amalgamation of CNN and LSTM networks' capabilities renders the proposed model especially adept for option price prognostications, given the pivotal role of historical prices and prevailing market dynamics in influencing future option valuations. Notably, the model manifests a precision rate of approximately 92% for option pricing over 200 epochs. Remarkably, even at a mere 20 epochs, the precision remains commendably high at around 90%. Hence, the data suggests that the ConvLSTM model can achieve substantial accuracy with a limited epoch count.

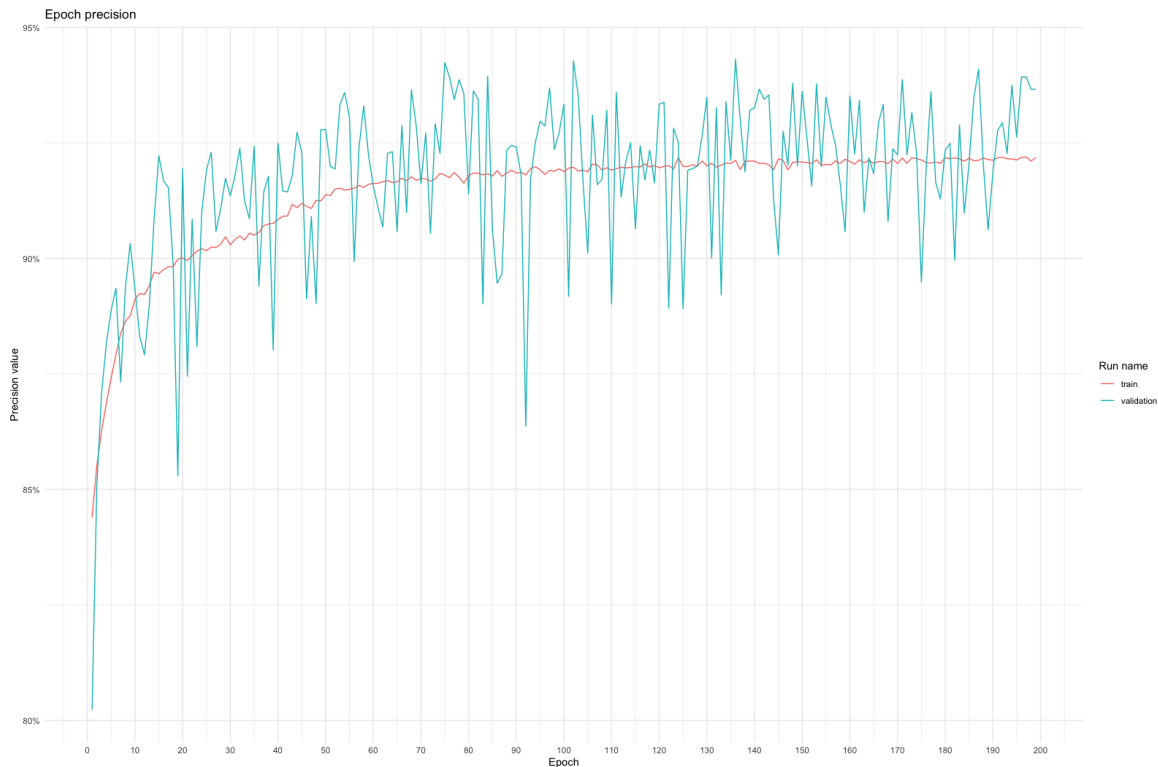


Fig. 2. Train and validation accuracy.

Within the ambit of financial predictive modeling, the Convolutional Long Short-Term Memory (ConvLSTM) model has emerged as a promising tool. Fig. 3 delves deep into a systematic examination of this model, particularly focusing on the intricacies of its training and validation losses across a span of 200 learning epochs. This graphical representation has been meticulously curated to elucidate the nuanced dynamics of the ConvLSTM model's learning process, offering a comprehensive view of its progressive enhancement in forecasting accuracy.

The primary intent behind Fig. 3 is manifold. First, it offers researchers, financial analysts, and readers an empirical visual assessment of how the model refines its predictive capabilities over successive epochs using the training dataset. By juxtaposing this with the model's performance on previously unseen validation data, the illustration provides a holistic perspective on the model's capacity to not only internalize data-driven patterns but also to generalize them effectively to new datasets. This aspect of generalization is pivotal in the real-world financial sphere, where the prediction model must navigate and adapt to constantly fluctuating market dynamics.

As one scrutinizes the trends depicted in Fig. 3, a clear attenuation in both training and validation losses becomes evident. This decline is not just indicative of the model's ability to reduce error margins over time, but it also underscores its robustness and adaptability in the face of evolving financial

data. It's noteworthy to mention that the sheer subtleness in the trajectory of these losses speaks volumes about the model's inherent stability and the efficiency of its learning algorithm.

Another salient observation from the figure is the relatively slight divergence between the training and validation losses. In the complex realm of deep learning, such congruence is often heralded as a testament to a model's consistent performance. A substantial gap could have implied potential overfitting, where the model becomes overly tailored to the training data and falters on new data. However, the narrow divergence observed reaffirms the ConvLSTM model's balanced approach, ensuring it retains its predictive accuracy even when confronted with unfamiliar financial datasets.

In summation, Fig. 3 stands as a pivotal piece of empirical evidence in this research. It not only attests to the ConvLSTM model's superior predictive capabilities but also illuminates its potential as a reliable tool for financial forecasting in a world characterized by uncertainty and rapid market fluctuations.

In the complex world of predictive modeling, understanding a model's accuracy and its capacity to generalize its learning is paramount. Fig. 4 serves as a compelling visualization of the Convolutional Long Short-term Memory (ConvLSTM) model's evolution in this regard, specifically focusing on the trajectory of its training and validation Root Mean Squared Error (RMSE) in the context of option price prediction [33].

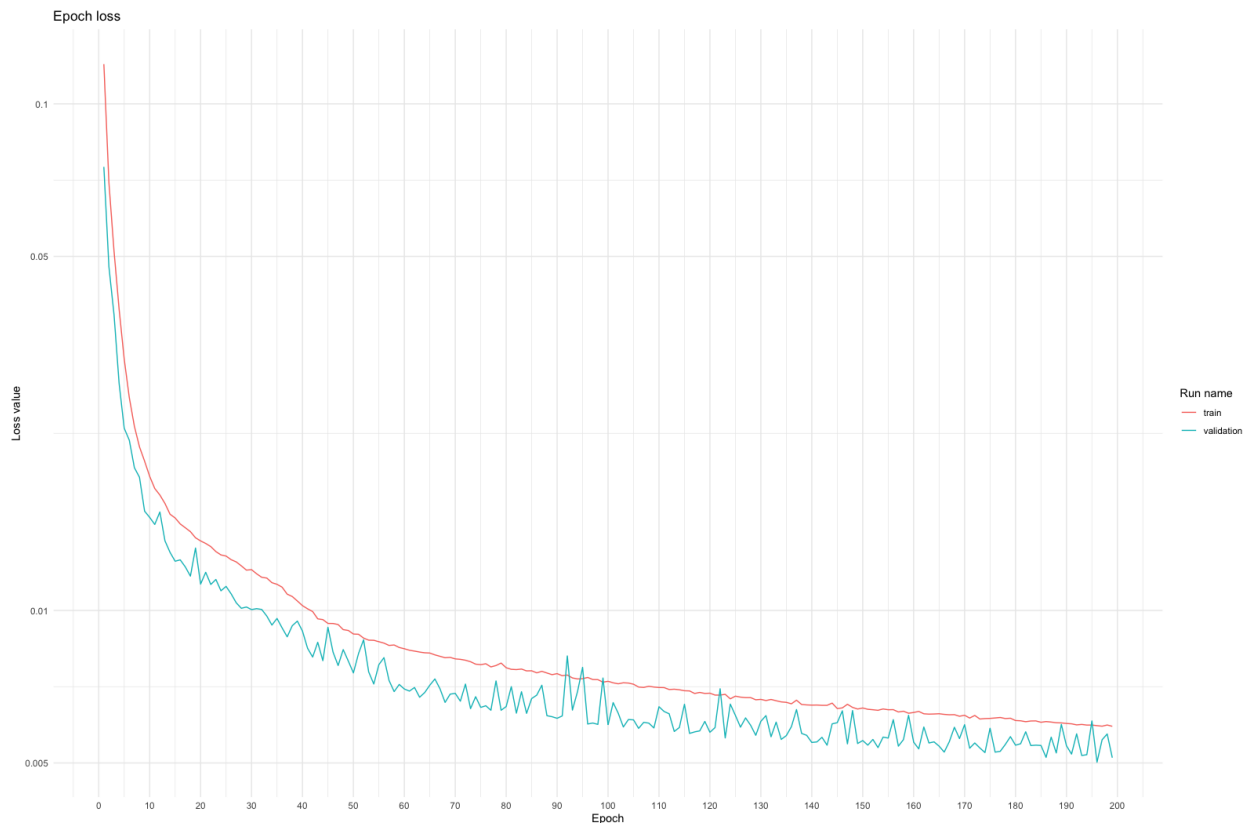


Fig. 3. Train and validation loss.

The figure meticulously chronicles the RMSE values over a delineated range of epochs. This choice of visualization enables a nuanced interpretation, offering readers an opportunity to juxtapose the model's learning curve (as evidenced by training RMSE) against its ability to adapt and predict new, unseen data (as evidenced by validation RMSE). Such a comparative study is vital in ascertaining the robustness and reliability of a predictive model, especially in dynamic domains like financial forecasting where precision is non-negotiable.

A foundational expectation in the realm of predictive analytics is that, as a model undergoes more training iterations (or epochs), its understanding of the data deepens, leading to enhanced prediction accuracy. This theoretical construct is manifestly demonstrated in Fig. 4. Both the training and validation RMSE values demonstrate a marked descent, which, in analytical parlance, signifies the ConvLSTM model's growing adeptness at decoding underlying data patterns and its prowess in translating this understanding into accurate forecasts.

The ConvLSTM architecture, hailed for its sophistication, has proven its mettle through the figure's depiction. Its RMSE values, even in initial epochs, are commendably modest, attesting to the architecture's inherent strength [34]. What further bolsters the ConvLSTM's credibility is the observed trend: as the model is exposed to more epochs, there's a palpable contraction in the RMSE. This diminishing RMSE trend, in conjunction with the relative proximity between

training and validation values, speaks volumes about the model's consistency and its adeptness at preventing overfitting.

In conclusion, Fig. 4 stands as a testament to the ConvLSTM model's exceptional capabilities in option price prediction. Through its clear delineation of RMSE progression, the figure elucidates the model's journey from initial understanding to mature precision, highlighting its potential as a dependable tool in the demanding sphere of financial analytics. Such insights underscore the significance of meticulous model evaluation and the promise that advanced architectures like ConvLSTM hold for future research endeavors.

Within the realm of financial forecasting, the veracity of a predictive model is often gauged by its ability to approximate real-world values. Fig. 5 serves as a visual testament to this critical evaluation, offering an intricate comparison between the ConvLSTM model's predicted option prices (rendered in blue) and the observed, actual option prices (depicted in red).

The visual portrayal in Fig. 5 invite a meticulous examination of the parallel trajectories of the forecasted and real option prices. The frequent confluence points between the blue and red lines provide compelling evidence of the model's predictive accuracy. Such intersections are not mere graphical intersections; they symbolize moments of congruence between predictive estimations and real-world observations. When a model's predictions consistently intersect with or closely trail the actual values, it is indicative of its robustness and precision.



Fig. 4. Root mean squared error changes.

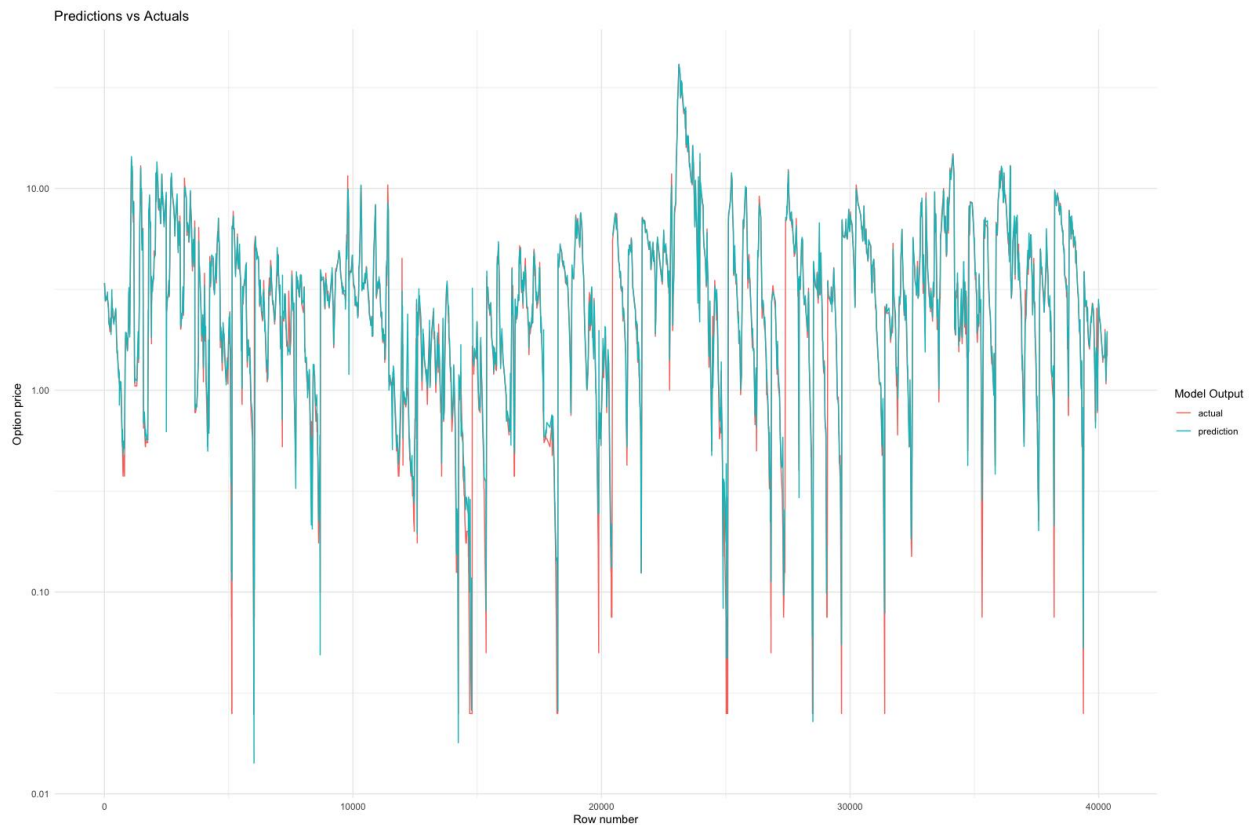


Fig. 5. Predicted and actual results for option pricing.

This observed alignment is not a trivial accomplishment, especially in the volatile domain of option pricing. The inherent unpredictability of financial markets makes the task of forecasting option prices a daunting challenge. Hence, the observed consistency between the model's estimations and the actual values underscores the ConvLSTM's superior modeling capabilities and its adeptness in capturing intricate market dynamics.

Moreover, the frequency of these intersections lends further credence to the ConvLSTM model's reliability. Infrequent matches could be dismissed as fortuitous, but the recurrent overlap observed in Fig. 5, suggests a pattern of sustained accuracy. Such consistency in predictions is emblematic of a model that has successfully internalized complex data patterns and can replicate this understanding in diverse scenarios.

Drawing from the insights garnered from Fig. 5, it becomes palpable that the ConvLSTM architecture's advanced design and computational prowess make it a formidable tool in the sphere of option price prediction. The close alignment between its forecasts and real prices is not merely a favorable outcome; it signifies the model's potential as a reliable instrument for long-term, practical applications in financial forecasting.

In conclusion, Fig. 5 stands as a testament to the ConvLSTM model's empirical efficacy. By providing a visual representation of the model's predictive prowess in the face of real-world data, the figure reinforces the notion that advanced predictive architectures like ConvLSTM are poised to revolutionize the landscape of financial analytics.

## V. DISCUSSION

The research undertaken here delves deep into the intricate facets of option price prediction using the Convolutional Long Short-term Memory (ConvLSTM) model, a hybrid deep learning architecture. The journey of analyzing, training, and testing the model has opened up avenues for reflections and discussions on both the capabilities of the model and the complexities associated with option pricing in financial markets.

Foremost, the accuracy metrics employed (e.g., RMSE, MAPE) offered crucial insights into the model's predictive performance. Notably, the ConvLSTM model showcased a capacity to capture both spatial and temporal aspects of the financial data. This is significant, as financial data streams inherently possess these dual characteristics. Traditional models often struggle with time-series data that also has spatial features, which makes ConvLSTM's accomplishment notable. The ability of the model to achieve around 92% precision in option pricing over 200 learning epochs stands as a testament to its robust nature. Further, its adeptness at reaching 90% precision in just 20 epochs underscores its efficiency.

The train and validation graphs over epochs – whether in terms of precision, loss, or RMSE – illuminated the model's learning curve. A significant observation was that the disparity between training and validation scores was minimal, hinting at minimal overfitting [35]. This implies that the model generalizes well to unseen data, a crucial factor in the volatile

world of financial markets where prediction robustness can translate to substantial economic implications.

The graphical representation comparing the predicted option prices with the actual values further reaffirmed the model's strength. The convergence of the blue and red lines not only indicates effective learning but also suggests the model's adaptability to evolving market dynamics.

Nevertheless, as with all models, it is essential to view these results in the broader context of financial modeling. The realm of option pricing is replete with volatility, influenced by an array of external factors such as geopolitical events, monetary policies, and global market trends [36]. While the ConvLSTM model has showcased commendable accuracy in this research, it is worth pondering how it might perform in extremely volatile scenarios or in the face of black swan events [37].

Another point of reflection is the comparison of the ConvLSTM model with traditional option pricing models like the Black-Scholes model [38]. While deep learning models like ConvLSTM offer adaptability and can work without strict assumptions, traditional models come with theoretical foundations deeply entrenched in financial theories. Would it be advantageous to integrate features from both worlds to achieve a more balanced, adaptable, yet theoretically sound model?

Data preprocessing, especially the normalization of parameters, was an essential step in this research. Neural networks, like the one employed in our ConvLSTM model [39], often require data to be structured and normalized to prevent issues like vanishing or exploding gradients [40]. It paves the path for a potential area of exploration: the development of models that are more resilient to raw or semi-processed data. Given the real-time nature of financial decisions, reducing preprocessing time without compromising accuracy could be invaluable.

Furthermore, the choice of data – specifically from the Chinese stock and fund market – brings forth questions about the model's adaptability to other global markets [41]. Financial behaviors and influences can vary across regions, affected by cultural, economic, and political differences. Hence, would the model retain its efficacy if trained on data from, say, the American or European markets?

In conclusion, the ConvLSTM model's potential in option price prediction, as evidenced by this research, is undeniably impressive. Its hybrid nature capitalizes on the strengths of both CNNs and LSTMs, making it a formidable tool for financial forecasting. However, it is imperative to continually evaluate and adapt the model, considering the ever-evolving landscape of global financial markets. The journey of this research serves not just as a testament to what has been achieved but also as an inspiration for the myriad possibilities that lie ahead.

## VI. CONCLUSION

The exploration into the predictive prowess of the Convolutional Long Short-term Memory (ConvLSTM) model for option price forecasting has yielded insightful revelations.

This research has undeniably demonstrated the potential of leveraging advanced deep learning architectures in the intricate and volatile realm of financial markets. As showcased, the ConvLSTM model adeptly captures the spatial and temporal nuances of financial data, achieving commendable accuracy levels over a minimal number of epochs.

A notable takeaway from this study is the synergy between the spatial feature detection capabilities of Convolutional Neural Networks (CNNs) and the sequential pattern recognition of Long Short-term Memory (LSTM) networks. The integration of these characteristics into the ConvLSTM model has proven to be particularly potent for forecasting in the dynamic domain of option pricing. The close alignment between the model's predictions and actual option prices underscores its viability as a tool for practical financial applications.

However, as with all predictive models, the ConvLSTM's performance should be perceived within the broader framework of financial analytics. While its adaptability and precision are commendable, continuous iterations and refinements are imperative, considering the multifaceted influences on financial markets. The model's ability to adapt to diverse global markets and unforeseen volatile events remains an avenue for future exploration.

In essence, this research has illuminated the vast potential that modern deep learning models hold for financial forecasting. It has set a benchmark, proving that with the right architecture and data, neural networks can be invaluable assets in the financial sector. As we advance, it is this confluence of finance and technology, exemplified by models like ConvLSTM that promises to redefine the contours of financial analytics and decision-making.

## REFERENCES

- [1] Wu, J., Wang, Z., Hu, Y., Tao, S., & Dong, J. (2023). Runoff forecasting using Convolutional Neural Networks and optimized bi-directional long short-term memory. *Water Resources Management*, 37(2), 937-953.
- [2] Wassie Geremew, G., & Ding, J. (2023). Elephant Flows Detection Using Deep Neural Network, Convolutional Neural Network, Long Short-Term Memory, and Autoencoder. *Journal of Computer Networks and Communications*, 2023.
- [3] Liu, C., Zhu, H., Tang, D., Nie, Q., Zhou, T., Wang, L., & Song, Y. (2022). Probing an intelligent predictive maintenance approach with deep learning and augmented reality for machine tools in IoT-enabled manufacturing. *Robotics and Computer-Integrated Manufacturing*, 77, 102357.
- [4] Tarek, Z., Shams, M. Y., Elshewey, A. M., El-kenawy, E. S. M., Ibrahim, A., Abdelhamid, A. A., & El-dosuky, M. A. (2023). Wind Power Prediction Based on Machine Learning and Deep Learning Models. *Computers, Materials & Continua*, 75(1).
- [5] 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.
- [6] 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.
- [7] Ajith, M., & Martínez-Ramón, M. (2023). Deep learning algorithms for very short term solar irradiance forecasting: A survey. *Renewable and Sustainable Energy Reviews*, 182, 113362.



- [8] 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.
- [9] Sun, Y., & Liao, K. (2023). A hybrid model for metro passengers flow prediction. *Systems Science & Control Engineering*, 11(1), 2191632.
- [10] Yin, L., & Wei, X. (2023). Integrated adversarial long short-term memory deep networks for reheater tube temperature forecasting of ultra-supercritical turbo-generators. *Applied Soft Computing*, 142, 110347.
- [11] Millham, R., Agbehadji, I. E., & Yang, H. (2021). Parameter tuning onto recurrent neural network and long short-term memory (RNN-LSTM) network for feature selection in classification of high-dimensional bioinformatics datasets. *Bio-inspired Algorithms for Data Streaming and Visualization, Big Data Management, and Fog Computing*, 21-42.
- [12] Vrskova, R., Sykora, P., Kamencay, P., Hudec, R., & Radil, R. (2021, July). Hyperparameter tuning of ConvLSTM network models. In *2021 44th International Conference on Telecommunications and Signal Processing (TSP)* (pp. 15-18). IEEE.
- [13] Mughees, N., Jaffery, M. H., Mughees, A., Mughees, A., & Ejsmont, K. (2023). Bi-LSTM-Based Deep Stacked Sequence-to-Sequence Autoencoder for Forecasting Solar Irradiation and Wind Speed. *Computers, Materials & Continua*, 75(3).
- [14] Durrani, A. U. R., Minallah, N., Aziz, N., Frnda, J., Khan, W., & Nedoma, J. (2023). Effect of hyper-parameters on the performance of ConvLSTM based deep neural network in crop classification. *Plos one*, 18(2), e0275653.
- [15] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436-444.
- [16] Yu, Y., Hu, G., Liu, C., Xiong, J., & Wu, Z. (2023). Prediction of solar irradiance one hour ahead based on quantum long short-term memory network. *IEEE Transactions on Quantum Engineering*.
- [17] Al-Alami, H., & Jamleh, H. O. (2023, May). Use of Convolutional Neural Networks and Long Short-Term Memory for Accurate Residential Energy Prediction. In *2023 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)* (pp. 294-299). IEEE.
- [18] Gosztolya, G., Busa-Fekete, R., Grósz, T., & Tóth, L. (2017). DNN-based feature extraction and classifier combination for child-directed speech, cold and snoring identification.
- [19] Viola, J., Chen, Y., & Wang, J. (2021). FaultFace: Deep convolutional generative adversarial network (DCGAN) based ball-bearing failure detection method. *Information Sciences*, 542, 195-211.
- [20] Dhiman, H. S., Deb, D., & Guerrero, J. M. (2022). On wavelet transform based Convolutional Neural Network and twin support vector regression for wind power ramp event prediction. *Sustainable Computing: Informatics and Systems*, 36, 100795.
- [21] Kumar, M., Singh, S., Kim, S., Anand, A., Pandey, S., Hasnain, S. M., ... & Deifalla, A. F. (2023). A Hybrid Model based on Convolution Neural Network and Long Short-Term Memory for Qualitative Assessment of Permeable and Porous Concrete. *Case Studies in Construction Materials*, e02254.
- [22] Zhou, C., Xu, H., Ding, L., Wei, L., & Zhou, Y. (2019). Dynamic prediction for attitude and position in shield tunneling: A deep learning method. *Automation in Construction*, 105, 102840.
- [23] Yarahmadi, A. M., Breuß, M., & Hartmann, C. (2022, September). Long Short-Term Memory Neural Network for Temperature Prediction in Laser Powder Bed Additive Manufacturing. In *Proceedings of SAI Intelligent Systems Conference* (pp. 119-132). Cham: Springer International Publishing.
- [24] Dalal, A. A., AlRassas, A. M., Al-qaness, M. A., Cai, Z., Aseeri, A. O., Abd Elaziz, M., & Ewees, A. A. (2023). TLIA: Time-series forecasting model using long short-term memory integrated with artificial neural networks for volatile energy markets. *Applied Energy*, 343, 121230.
- [25] Garg, S., & Krishnamurthi, R. (2023). A survey of long short term memory and its associated models in sustainable wind energy predictive analytics. *Artificial Intelligence Review*, 1-50.
- [26] Alamri, N. M. H., Packianather, M., & Bigot, S. (2023). Optimizing the parameters of long short-term memory networks using the bees algorithm. *Applied Sciences*, 13(4), 2536.
- [27] Fati, S. M., Muneer, A., Alwadain, A., & Balogun, A. O. (2023). Cyberbullying Detection on Twitter Using Deep Learning-Based Attention Mechanisms and Continuous Bag of Words Feature Extraction. *Mathematics*, 11(16), 3567.
- [28] Mubashar, M., Khan, G. M., & Khan, R. (2021, April). Landslide prediction using long short term memory (LSTM) neural network on time series data in Pakistan. In *2021 International conference on artificial intelligence (ICAI)* (pp. 175-181). IEEE.
- [29] Becerra-Rico, J., Aceves-Fernández, M. A., Esquivel-Escalante, K., & Pedraza-Ortega, J. C. (2020). Airborne particle pollution predictive model using Gated Recurrent Unit (GRU) deep neural networks. *Earth Science Informatics*, 13, 821-834.
- [30] Yessoufou, F., & Zhu, J. (2023). Classification and regression-based Convolutional Neural Network and long short-term memory configuration for bridge damage identification using long-term monitoring vibration data. *Structural Health Monitoring*, 14759217231161811.
- [31] Panda, B., & Singh, P. (2023). A deep convolutional-LSTM neural network for signal detection of downlink NOMA system. *AEU-International Journal of Electronics and Communications*, 154797.
- [32] Chaudhury, S., & Sau, K. (2023). A BERT encoding with Recurrent Neural Network and Long-Short Term Memory for breast cancer image classification. *Decision Analytics Journal*, 6, 100177.
- [33] Liu, H., Yan, G., Duan, Z., & Chen, C. (2021). Intelligent modeling strategies for forecasting air quality time series: A review. *Applied Soft Computing*, 102, 106957.
- [34] Girsang, A. S., & Tanjung, D. (2022). Fast Genetic Algorithm for Long Short-Term Memory Optimization. *Engineering Letters*, 30(2).
- [35] Conti, P., Guo, M., Manzoni, A., & Hesthaven, J. S. (2023). Multi-fidelity surrogate modeling using long short-term memory networks. *Computer methods in applied mechanics and engineering*, 404, 115811.
- [36] Zhang, C., Pan, G., & Fu, L. (2023). Real-Time Intersection Turning Movement Flow Forecasting Using a Parallel Bidirectional Long Short-Term Memory Neural Network Model. *Transportation Research Record*, 03611981231172958.
- [37] Lu, W., Duan, J., Wang, P., Ma, W., & Fang, S. (2023). Short-term wind power forecasting using the hybrid model of improved variational mode decomposition and maximum mixture correntropy long short-term memory neural network. *International Journal of Electrical Power & Energy Systems*, 144, 108552.
- [38] Stankovic, M., Bacanin, N., Budimirovic, N., Zivkovic, M., Sarac, M., & Strumberger, I. (2023, April). Bi-Directional Long Short-Term Memory Optimization by Improved Teaching-Learning Based Algorithm for Univariate Gold Price Forecasting. In *2023 International Conference on Inventive Computation Technologies (ICICT)* (pp. 1650-1657). IEEE.
- [39] Munsif, M., Ullah, M., Fath, U., Khan, S. U., Khan, N., & Baik, S. W. (2023). CT-NET: A Novel Convolutional Transformer-Based Network for Short-Term Solar Energy Forecasting Using Climatic Information. *Computer Systems Science & Engineering*, 47(2).
- [40] Dong, S., Wang, P., & Abbas, K. (2021). A survey on deep learning and its applications. *Computer Science Review*, 40, 100379.
- [41] Rama-Maneiro, E., Vidal, J., & Lama, M. (2021). Deep learning for predictive business process monitoring: Review and benchmark. *IEEE Transactions on Services Computing*.

# Factors and Models Influencing Value Co-Creation in the Supply Chain of Collection Resources for Library Distribution Providers Under Data Ecology

Xiaoyun Lin\*

Library

Guangzhou College of Technology and Business  
Guangzhou, 510850, China

**Abstract**—Under the data ecology, the advancement of relevant technology and the utilization of relevant resources have provided more efficient technical services for various industries. However, with the proliferation of data resources, problems such as information pollution and data redundancy have arisen in the process of supply chain services for collection resources. For solving such problems and enhancing the collection resource supply efficiency of librarians, the study uses data mining technology combined with improved K-Means clustering algorithm to design a value co-creation model of library collection resource supply chain for librarians under data ecology. The outcomes indicate that the shortest running time of traditional K-Means algorithm is 40ms and the longest running time is 115ms in Wine dataset, and the running time of improved K-Means algorithm is stable at 59ms; the shortest running time of traditional K-Means algorithm is 26 ms and the longest running time is 58ms in Iris dataset, and the running time of improved K-Means algorithm is stable at 53 ms. The clustering accuracy of the improved K-Means algorithm in the Wine data set is 98.2%, which is 0.3% exceeding the traditional K-Means algorithm, which is 97.9%; the clustering accuracy in the Iris data set is 100%, which is 2.4% exceeding the traditional K-Means algorithm, which is 97.6%. In summary, it can be seen that the studied data ecology has a good application of the factors and models influencing the value co-creation of the supply chain of library resources for library dispensers.

**Keywords**—Resource supply chain; data mining; value co-creation; K-Means clustering algorithm; pavilion dispenser

## I. INTRODUCTION

With the big data's reach, the use of data resources changed the current human lifestyle and cognitive level, and data resources have become an important asset [1]. The advancement of relevant technology has led to the increasing participation of users in the network, and the phenomenon that everyone has data but everyone lacks data has emerged in various industries [2]. The increase in data transparency and freedom and the increasing circulation of data have led to information pollution and data redundancy in the current data ecosystem [3]. Therefore, the traditional library resource management model cannot satisfy the needs of efficient resource integration for the massive library collection resources nowadays. Data mining (DM) is a data-centered mode of thinking, which can rely on algorithms to effectively refine and integrate generous data to maximize the benefits of

data resources [4]. The integration of resources based on DM technology is the service goal of the supply chain of collection resources of the librarians, and the deep combination of DM and supply chain resource management to explore a high-efficiency and low-cost resource supply chain management model is the method for realizing the relevant value of collection resources. The K-Means (KM) algorithm is extensively utilized classical classification clustering algorithms (CA). It can effectively solve the low clustering accuracy and simple for falling into local solutions [5]. The problem and goal of this article are to improve the efficiency of collection resource supply for library distributors. The significance and contribution of this study are to propose the connotation of value co-creation (VCC) in the collection resource supply chain of library distributors, provide readers with a better knowledge service environment, and provide theoretical basis and direction guidance for future research on the collection resource supply chain. Therefore, in this context, the factors and models influencing the VCC in the supply chain of library resources for library distributors under the data ecology are studied to improve the use rate and realize the VCC in the supply chain of library resources. The second part is a review of the research of resource supply chain and DM technology at home and abroad; the third part is a study on the VCC model of library resource supply chain based on the data ecology. The first section analyzes the factors affecting the VCC of library resource supply chain based on the Decision-making Trial and Evaluation Laboratory (DEMATEL)\_ method; the second section is the constructing of the VCC model of library resource supply chain for library distributors. The fourth part is the validation of the VCC model of the supply chain of library resources for library service providers based on the data ecology.

## II. RELATED WORKS

The resource distribution link of the resource supply chain introduces new-age technologies such as logistics technology and Internet of Things. Therefore, the resource supply chain based on data and emerging technologies has gained the attention of many domestic and foreign experts and scholars, and researchers have carried related research on it. Nandi M L and other scholars proposed integrating relevant technology into its corresponding supply chain system to improve the supply chain performance and construct a resource-based

supply chain system framework. It is based on 126 secondary information cases and a qualitative content analysis of these 126 supply chain system cases using a retrospective approach. The outcomes illustrated that the method is useful in improving information sharing in supply chain and has reference value in supply chain integration [6]. The research team of Agyabeng-Mensah Y, to test the influence of green human resource management on internal green supply chain, proposed an equation modeling software with partial least squares structure, which utilized a relevant method to collect data in supply chain and used partial least squares structure for analyzing the data. The outcomes showed that the method tested the internal green supply chain and was feasible [7]. Mehrotra S et al. proposed to establish a stochastic optimal allocation model with sharing function for enhancing the system of the ventilator supply chain, which was mainly used for the allocation of ventilator resources during the new coronary pneumonia pandemic. The outcomes indicated it could enhance the allocation efficiency of the ventilator supply chain [8].

DM techniques possess an essential influence in resource supply chain research. Zhao Y's research team proposed a building energy fault diagnosis model based on DM techniques for enhancing the operation of energy systems in the building industry. Then the model utilized unsupervised DM techniques to construct a framework for building energy operation pattern identification and energy load diagnosis, and the model was tested in a dataset Validation. The outcomes indicated that the model possessed excellent accuracy of building energy load prediction [9]. Liu J and other scholars proposed a privacy-preserving DM algorithm in view of Bayesian analysis and logical analysis for addressing the data privacy protection. The core of the method was for letting DM technology become data and strengthen the management of DM technology. The outcomes indicated that the method improved the security of privacy protection by 20% and is feasible [10]. Rong Z and other researchers proposed a multi-layer fuzzy evaluation model in view of AI DM, which integrated fuzzy information such as learning, recognition, correlation and adaptive to evaluate the ideological education of undergraduates, in response to the problems of unclear evaluation system and single evaluation method of

undergraduates. The outcomes illustrated that the model achieved the desired effect with scientific validity [11].

In summary, DM techniques are often applied in supply chain research in various fields, but research on using big DM techniques to build a VCC model for the supply chain of library resources is quite rare. In this context, the factors and models influencing the VCC of the supply chain of collection resources for library dispensers under the data ecology are studied to achieve better application results.

### III. RESEARCH ON THE VCC OF THE SUPPLY CHAIN OF LIBRARY RESOURCES IN VIEW OF THE DATA ECOLOGY OF LIBRARY DISPENSERS

In this chapter, the structure of the supply chain of library resources of library dispensers under the data ecology is studied, and the factors influencing the VCC of the supply chain of library resources are analyzed using the DEMATEL method, and the construction of the VCC model of the supply chain of library resources of library dispensers is carried out using the improved KM CA.

#### A. Analysis of the Factors Influencing the VCC in the Supply Chain of Collection Resources based on the DEMATEL Method

Data ecology is evolved from the digitization of knowledge resources, which is generated as an outcome of the combination of resource digitization and information technology (IT). Data ecology is based on big data carriers, forming a community of interest for data sharing and collaboration to achieve the purpose of VCC [12]. From the perspective of supply chain, libraries can be regarded as logistics distribution centers, and the resource supply chain links upstream collection suppliers and downstream library customers. The theory of VCC in the supply chain of collection resources can be used in the library distribution industry, where libraries and distribution providers work together to provide quality collection resources to library customers and create social value to improve social and economic benefits [13]. The supply chain structure of library resources for library distributors under data ecology is shown in Fig. 1.

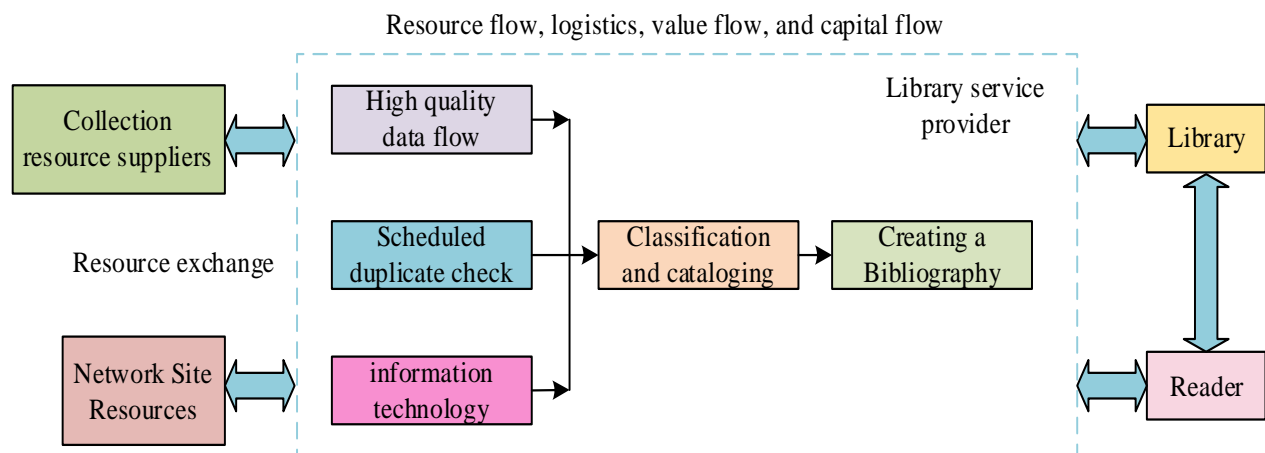


Fig. 1. The supply chain structure of collection resources for library distributors in the data ecology.

As shown in Fig. 1, the major part of the supply chain structure of collection resources mainly contains libraries, library distributors, library customers and collection suppliers. The collection resources mainly include book resources, e-book resources and library intelligent auxiliary devices provided by the collection suppliers. To realize the VCC in the supply chain of collection resources, it is essential for bringing into play the supply efficiency of collection resources and ensures the effective circulation of resource flow, capital flow, logistics and value flow in the supply chain of collection resources [14]. According to the root theory, nine factors can be derived to influence the supply chain VCC: the construction and integration of supply chain resources, the advancement and applying of supply chain service technology, the applying of data for supply chain VCC, the quality of staff in the supply chain, the concept of supply chain VCC, the construction of supply chain VCC platform, the interaction of the main body of supply chain VCC, the demand of readers for supply chain VCC, and the environmental constraints of supply chain VCC, etc. [15]. DEMATEL is a way of systematic analysis using matrix tools and graph theory, which can calculate the influence degree of each factor based on the logical relation in the elements [16]. For identifying the key factors in the supply chain VCC from the nine factors, the degree of correlation between the factors is calculated using the DEMATEL method. The mathematical expression of the normalization matrix is shown in Eq. (1).

$$H = \frac{X}{\max_{1 \leq i \leq n} \sum_{j=1}^n x_{ij}} \quad (1)$$

In Eq. (1),  $X$  denotes the influence relationship matrix while  $H$  means the normalized one.  $x_{ij}$  denotes the degree of influence of factor  $i$  on factor  $j$ . The relevant formula is demonstrated in Eq. (2).

$$S = \lim_{k \rightarrow \infty} (H + H^2 + \dots + H^k) = \sum_{k=1}^{\infty} H^k = H(K - H)^{-1} \quad (2)$$

In Eq. (2),  $K$  denotes the unit matrix and  $S$  indicates the combined influence matrix. The method for calculating the influence degree of each factor is shown in Eq. (3).

$$Q = \sum_{j=1}^n s_{ij} \quad (i = 1, 2, \dots, N) \quad (3)$$

In Eq. (3),  $Q$  denotes the influence degree vector of factors and  $N$  expresses the total of factors. The influenced degree vector is shown in (4).

$$W = \sum_{i=1}^n s_{ij} \quad (j = 1, 2, \dots, N) \quad (4)$$

In Eq. (4),  $W$  indicates the vector of factors' influence degree, and the centrality and the reason degree can be calculated according to the influence degree of the factors. Reason degree illustrates the influence of the factor on other factors, and a positive reason degree illustrates that the element possesses an impact on other elements, while a negative reason degree illustrates that the element is affected by other elements [17]. The relevant formula is showcased in Eq. (5).

$$U_i = Q_i + W_i \quad (5)$$

In Eq. (5),  $U$  indicates the value of centrality. The relevant formula is showcased in Eq. (6)

$$V_i = Q_i - W_i \quad (6)$$

In Eq. (6),  $V$  indicates the value of the cause degree. The combined importance of the system elements can be counted from the values of centrality and causality, as shown in Eq. (7).

$$Z_i = \alpha U_i + \beta V_i \quad (7)$$

In Eq. (7),  $Z_i$  represents the combined importance of system factors;  $\alpha$  denotes the pendulum weighting coefficient of centrality, and  $\beta$  refers to the pendulum weighting coefficient of cause. Based on the maximum and minimum values of the centrality and the reason degree, the optimal and the worst solutions can be derived, and then the swing weight coefficients are used to assign weights to the indicators of the two solutions. Then the final weight values of the centrality and the reason degree can be obtained by normalizing the two assigned weights. Based on the key factors identified by the DEMATEL method, a model of co-creation influence factors of the supply chain value of the collection resources based on the DEMATEL method is designed as shown in Fig. 2.

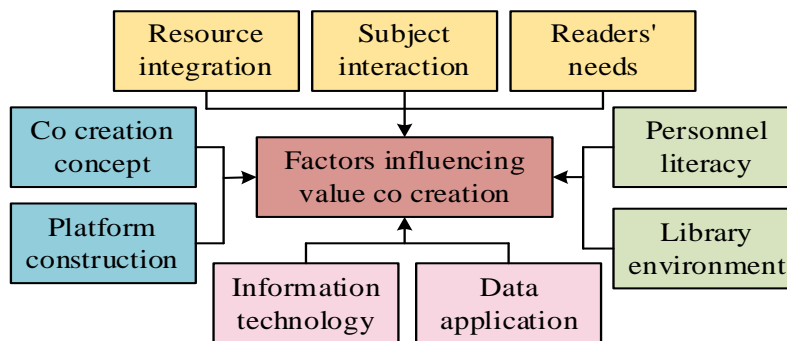


Fig. 2. A model of factors influencing the VCC of collection resources supply chain based on DEMATEL method.

As shown in Fig. 2, in the VCC influence factor model, platform construction and co-creation concept are the key factors of supply chain VCC of library resources; readers' demand, subject interaction and resource integration are the drivers of supply chain VCC; platform environment and staff literacy are the direct factors of supply chain VCC; and data applying and IT are the indirect factors of supply chain VCC.

**B. The Construction of the VCC Model of the Supply Chain of Library Resources for Library Distribution Providers**

The realizing of VCC in the supply chain of collection resources under the data ecology cannot be achieved without the influence of nine factors, while the complexity of the operation and management of the supply chain of collection resources brings challenges to the realization of VCC. For reaching the purpose of providing readers with quality resources in the resource supply chain, DM technology is utilized for exploring the VCC in the collection resource supply chain and build a model of VCC in the collection resource supply chain with DM. DM is a data-centered mode

of thinking, and DM relies on algorithms and technologies to effectively refine and integrate massive amounts of data to maximize the benefits of data resources [18]. For achieving the VCC process of the supply chain of collection resources under the data ecology, the VCC model of the supply chain of collection resources for library distribution providers is constructed as shown in Fig. 3.

As shown in Fig. 3, the data base in the VCC model of the supply chain of library resources contains equipment operation data, organization operation data, reader behavior data and resource integration data. Effective DM of the collection resources under the big data ecology is beneficial for the library distributor to understand the information of readers' needs and library business weaknesses, and it can also enhance the communication efficiency among the supply chain subjects. The core idea of DM technology is mining potentially useful information from generous disordered and incomplete data using relevant algorithms. The relevant basic process is showcased in Fig. 4.

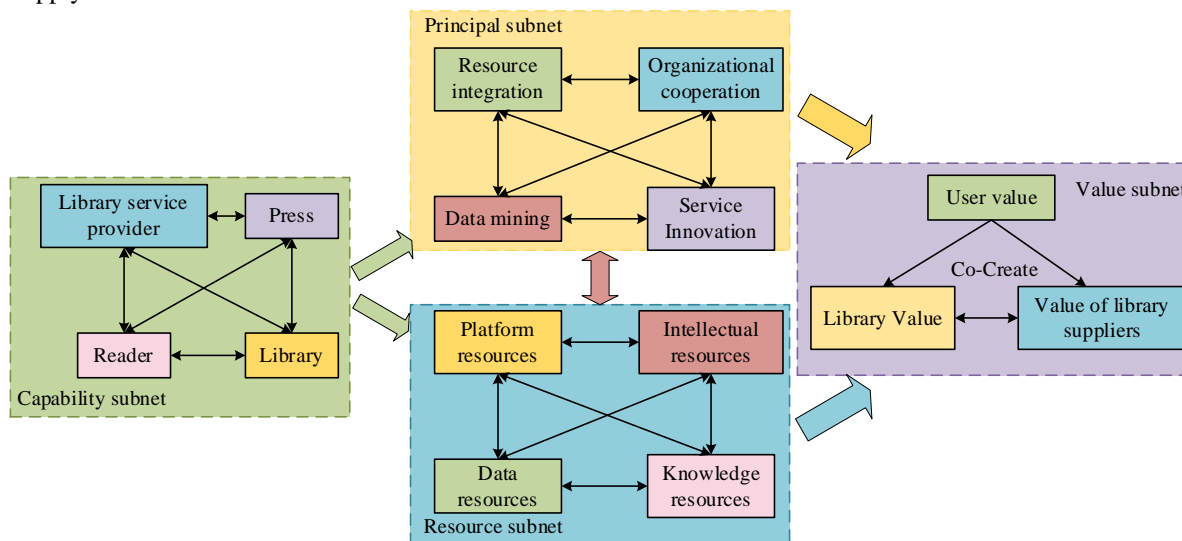


Fig. 3. The VCC model of collection resources supply chain for library distributors.

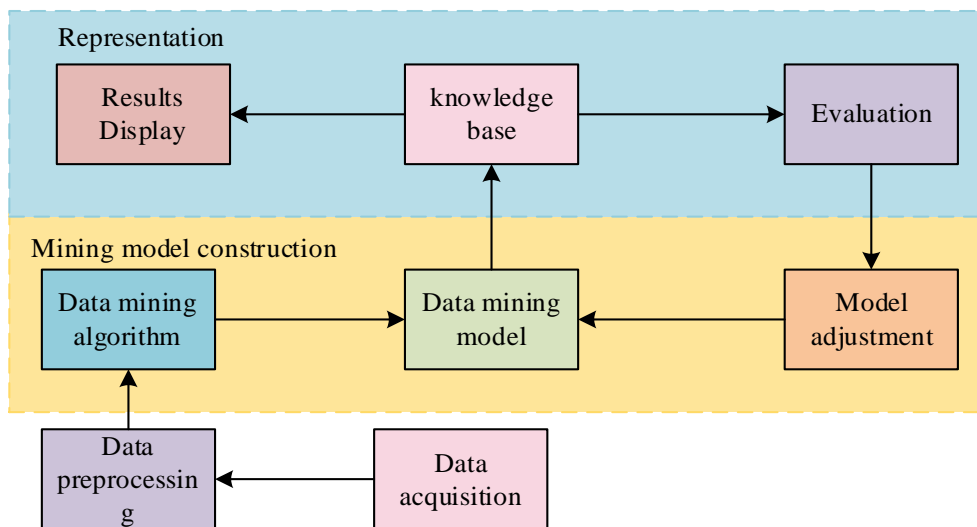


Fig. 4. The basic process of DM.

Fig. 4 indicates that the DM includes steps of data collection, data pre-processing, construction of DM models and output of results. The algorithms for building DM models are generally divided into four categories, such as CA, association rule algorithms, classification and regression algorithms, and CA [19]. The KM algorithm is extensively utilized CA as its simple principle and easy implementation [20]. The CA is an indirect clustering division method, which can measure the similarity between samples using Euclidean distance (ED) and transform it by means of data distance metric for better calculation of data. The CA utilizes distance as a criterion for evaluating the metric of sample similarity. The closer the distance in two samples, the higher the similarity between the samples. The most used distance formulas in CA are ED and Manhattan distance, and the mathematical expression of ED is indicated in Eq. (8).

$$D(i, j) = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{in} - x_{jn})^2} \quad (8)$$

In Eq. (8),  $D$  denotes the ED,  $i$  and  $j$  denote the data objects, and  $n$  denotes the dimensionality. The mathematical expression of the Manhattan distance is shown in Eq. (9).

$$D'(i, j) = |x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + \dots + |x_{in} - x_{jn}| \quad (9)$$

In Eq. (9),  $D'$  denotes the Manhattan distance. If no end criterion is set in the KMA, the distance calculation of Eq. (8) and Eq. (9) will be repeated, and reaching the end mark means that the algorithm has reached convergence. The flow chart of the KMA is illustrated in Fig. 5.

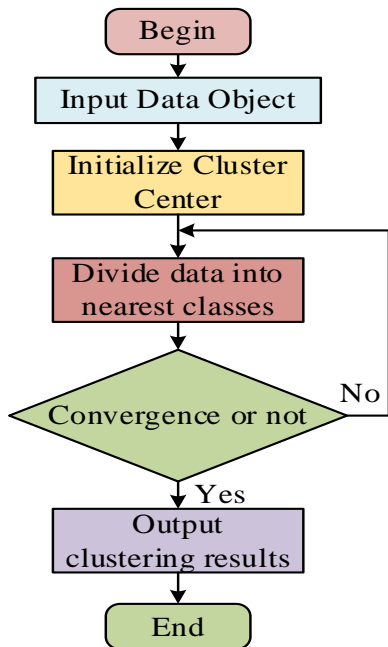


Fig. 5. Flowchart of KM algorithm.

Fig. 5 indicates that the algorithm first prepares the data, determines the number of classifications and sets the clustering centers, then counts the range in each data point and the clustering center, and groups the closest points into the

initial centroids. After that, the clustering centroids are recalculated and the convergence is judged. If the convergence is not reached, the above steps are repeated until convergence is achieved. The most common method used in CA to determine whether convergence is achieved is the error-sum-squared criterion method, and the mathematical expression of the error-sum-squared criterion method is shown in Eq. (10).

$$J_D = \sum_{i=1}^D \sum_{k=1}^{m_i} \|x_k - m_i\|^2 \quad (10)$$

In Eq. (10),  $D$  denotes the sample set;  $J$  denotes the objective function;  $m_i$  serves as the mean value of the sample, and the formula for calculating the sample mean is shown in Eq. (11).

$$m_i = \frac{1}{n_i} \sum_{i=1}^{n_i} x_i \quad i = 1, 2, \dots, D \quad (11)$$

According to Eq. (10) and Eq. (11), the value of the objective function depends on the sample centroid, and the larger the sample centroid, the larger the objective function and consequently the larger the error; conversely, the smaller the sample centroid, the smaller the objective function, and consequently the smaller the error. For reducing the probability of error in the calculation, the traditional KM algorithm is enhanced by using the weighted squared average distance method to reduce the search range in the calculation. The expression of the weighted squared average distance method is shown in Eq. (12).

$$J_i = \sum_{i=1}^D P_i A_i^* \quad (12)$$

In Eq. (12),  $A_i^*$  denotes the mean squared distance between samples and  $P_i$  denotes the prior probability. The formula for calculating the mean squared distance is shown in Eq. (13).

$$A_i^* = \frac{2 \sum_{x \in X_i} \sum_{x' \in X_i} \|x - x'\|^2}{n_i (n_i - 1)} \quad (13)$$

In Eq. (13),  $n_i$  denotes the number of data samples, and  $\sum_{x \in X_i} \sum_{x' \in X_i} \|x - x'\|^2$  denotes the sum of distances between samples. The mathematical expression of the interclass distance and criterion is shown in Eq. (14).

$$J_{\delta 1} = \sum_{i=1}^D (m_i - m)^T (m_i - m) \quad (14)$$

In Eq. (14),  $m$  denotes the mean value of the distance between samples, and  $m_i$  denotes the mean value of the distance between samples of  $i$ . The weighted interclass distances and criteria are shown in Eq. (15).

$$J_{\delta 2} = \sum_{i=1}^D P_i (m_i - m)^T (m_i - m) \quad (15)$$

As shown by Eq. (15), the weighted interclass distance and criterion is the association of adding a priori probability to the interclass distance and criterion formula. The larger the value of  $J_{\delta}$  corresponds to the greater the degree of variation, then the more obvious the clustering analysis results.

#### IV. VALIDATION OF THE VCC OF THE SUPPLY CHAIN OF LIBRARY RESOURCES IN VIEW OF THE DATA ECOLOGY OF LIBRARY DISTRIBUTORS

For verifying the performance of the VCC model of the supply chain of library resources based on the data ecology, the performance of the traditional KM algorithm was compared with that of the improved one. The total samples in the Wine dataset were 440, and there were four clusters of the same size and no isolated points in the dataset; the total number of samples in the Iris dataset was 300, and there were three clusters of the same size and five isolated points in the dataset. The specific the relevant outcomes are showcased in Table I.

The traditional and the improved KM algorithms were run ten times in each of the two training sets, and the comparing of the CA of the two algorithms run ten times in the data set is shown in Fig. 6. Fig. 6(a) illustrated that the traditional KM

algorithm showed an up-and-down state in the Wine dataset, with the highest CA of 70.3% and the lowest of 53.5%, while the CA of the improved KM algorithm was stable at 71%. In Fig. 6(b), the traditional KM algorithm achieved the lowest CA of 65% at the sixth run in the Iris dataset, and the CA fluctuated around 89% for the other runs, while the CA of the improved KM algorithm remained at 90%. Collectively, it illustrated that the traditional KM algorithm was not stable enough and the improved KM algorithm had more stability.

For comparing the performance of the traditional and the improved KM algorithms more comprehensively, the running times of the two algorithms were studied. The running time comparison of the two algorithms run ten times in the dataset is indicated in Fig. 7. Fig. 7(a) showcased that the running time of the traditional KM algorithm in the Wine dataset still presented an unstable state, in which the shortest running time was 40 ms and the longest running time was 115 ms, while the running time of the improved KM algorithm was stable at 59 ms. Fig. 7(b) indicated that the traditional KM algorithm in the Iris dataset had the shortest running time of 40 ms and the longest running time of 115 ms, while the running time of the improved KM algorithm was stable at 59 ms. The shortest running time of the traditional KM algorithm was 26ms and the longest running time was 58ms, while the running time of the improved KM algorithm stayed at 53ms. Collectively, it illustrated that the improved KM algorithm had stronger robustness compared to the traditional KM algorithm.

TABLE I. EXPERIMENTAL ENVIRONMENT

Experimental Environment	Configuration
Operating system	Windows 10
Memory	16G
GPU	NVIDIA TITAN BLACK GPU
CPU	Intel(R) Core (TM) i5-4460
Programming Language	Python
Algorithm testing tools	PyCharm

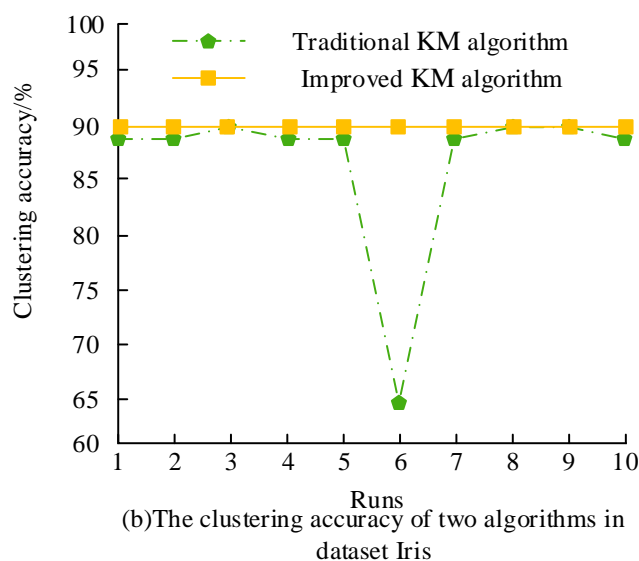
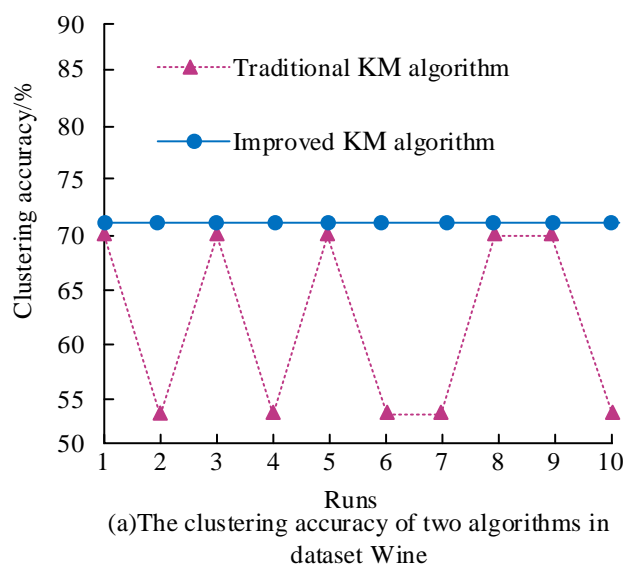


Fig. 6. Comparison of CA between two algorithms running ten times.

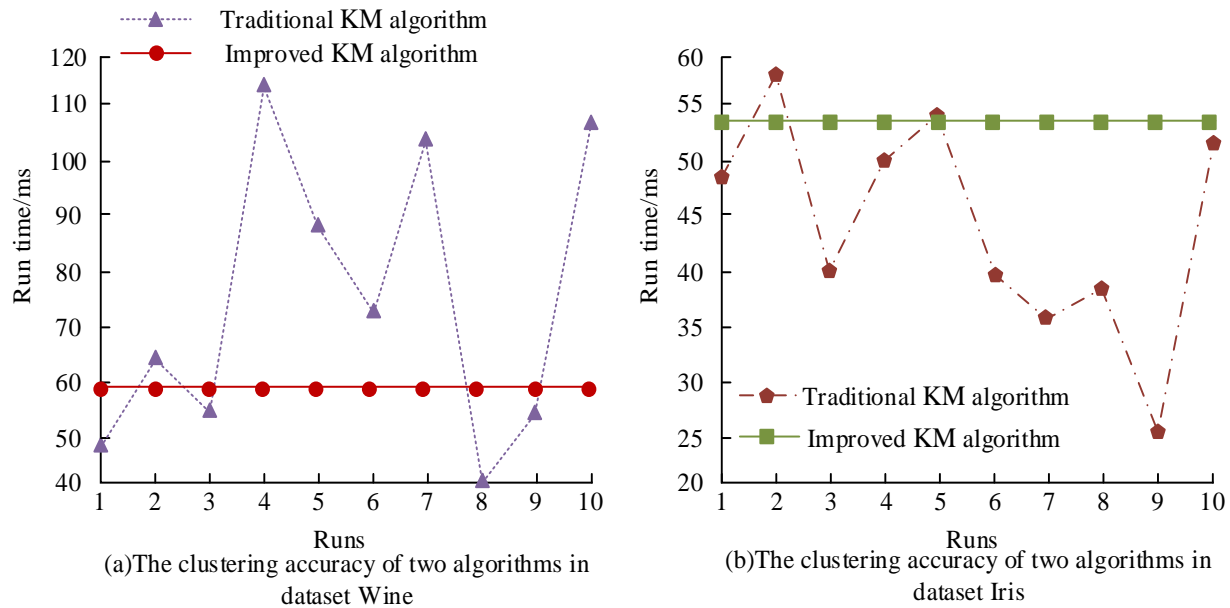


Fig. 7. Comparison of the runtime of two algorithms in the dataset.

To observe more intuitively the performance comparing between the traditional and the improved KM algorithms, the performance metrics of the two algorithms were compared and analyzed with the number of iterations. The comparison results are showcased in Table II. From Table II, the average CA of the improved KM algorithm was 71% and 90% in the Wine dataset and Iris dataset, respectively, which was 9.1% and 3.4% higher than the average CA of the traditional KM algorithm. And it showcased that the improved KM algorithm had higher CA, and the improved KM algorithm only runs the best clustering effect could be achieved by running only once. In terms of the number of iterations, that of the improved KM algorithm was smaller than the average iteration times of the traditional KM algorithm. Although the average running time of the improved KM algorithm was 53 ms in the Iris dataset, which was 8.6 ms longer than the average running time of 44.4 ms of the traditional KM algorithm, the improved KM algorithm could achieve the best clustering effect after only

one run. Therefore, the improved KM algorithm still achieved a more desirable performance in terms of running time.

From verifying the traditional and the improved KM algorithms in practical applications, the two algorithms are compared and analyzed in simulation experiments in the Wine dataset. The outcomes of the simulation experimental runs of the two algorithms in the Wine dataset are shown in Fig. 8. Fig. 8(a) indicated that the clustering result of the Wine dataset contained four clusters of almost the same size and no isolated points in the dataset. Fig. 8(b) illustrated that the clustering results of the traditional KM algorithm in the Wine dataset produced misclassification rates of 1.2%, 3.5%, 1.3% and 2.5% for cluster 1, cluster 2, cluster 3 and cluster 4, respectively, and the CA was 97.9%. As seen in Fig. 8(c), the clustering results of the improved KMA in the Wine dataset yielded misclassification rates of 1.3%, 3.5%, 0 and 2.1% for cluster 1, cluster 2, cluster 3 and cluster 4, with a CA of 98.2%.

TABLE II. COMPARISON OF PERFORMANCE INDICATORS BETWEEN TRADITIONAL AND IMPROVED KM ALGORITHMS

Performance Index	Data Set	Traditional KM Algorithm			Improved KM Algorithm		
		Minimum	Highest value	Average	Minimum	Highest value	Average
Iterations	Wine	4	12	8.3	7	7	7
	Iris	5	13	9	8	8	8
Clustering accuracy/%	Wine	53.5	70.3	61.9	71	71	71
	Iris	65	90	86.6	90	90	90
Run time/ms	Wine	40	115	75.4	59	59	59
	Iris	26	58	44.4	53	53	53



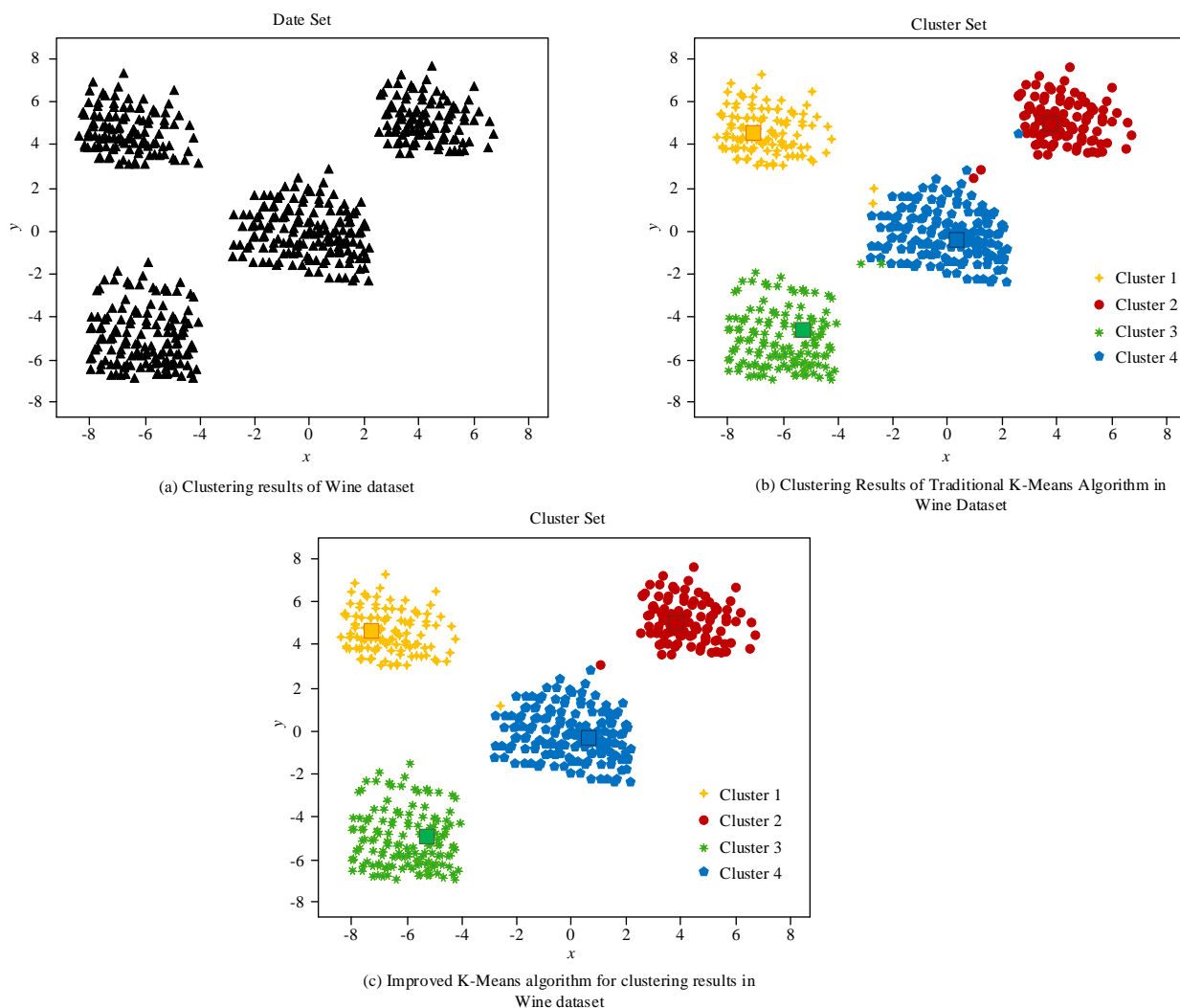


Fig. 8. Simulation experimental results of two algorithms in Wine dataset.

For more comprehensively verifying the traditional and the improved KM algorithms in practical applications, the two algorithms were compared and analyzed in simulation experiments on the Iris dataset. The results of the simulation experimental runs of the two algorithms in the Iris dataset are shown in Fig. 9. Fig. 9(a) demonstrated that the clustering outcome of the Iris dataset contained three clusters of almost the same size and there were five isolated points in the dataset. In Fig. 9(b), the traditional KM algorithm's clustering outcomes on the Iris dataset yielded misclassification rates of 2%, 1.1%, and 4% for clusters 1, 2, and 3, respectively, with a CA of 97.6%. Fig. 9(c) showcased the clustering outcomes of the improved KM algorithm in the Iris dataset, cluster 1, cluster 2, and cluster 3 all produced a misclassification rate of 0, respectively, with a CA of 100%.

For more intuitively observing the comparison of effectiveness of the traditional and the improved KM algorithms in practical applications, the clustering results in the two datasets were evaluated. The clustering evaluations of the traditional and the improved KM algorithms are shown in Table III. Table III illustrated that the CA of the improved KM algorithm in the Wine dataset was 98.2%, which was 0.3% higher than the CA of 97.9% of the traditional KM algorithm. In the Iris dataset, the CA of the improved KM algorithm reached 100%, which was 2.4% higher than the CA of 97.6% of the traditional KM algorithm. Collectively, the CA of the improved KM algorithm significantly exceeded that of the traditional KM algorithm in practical applications.

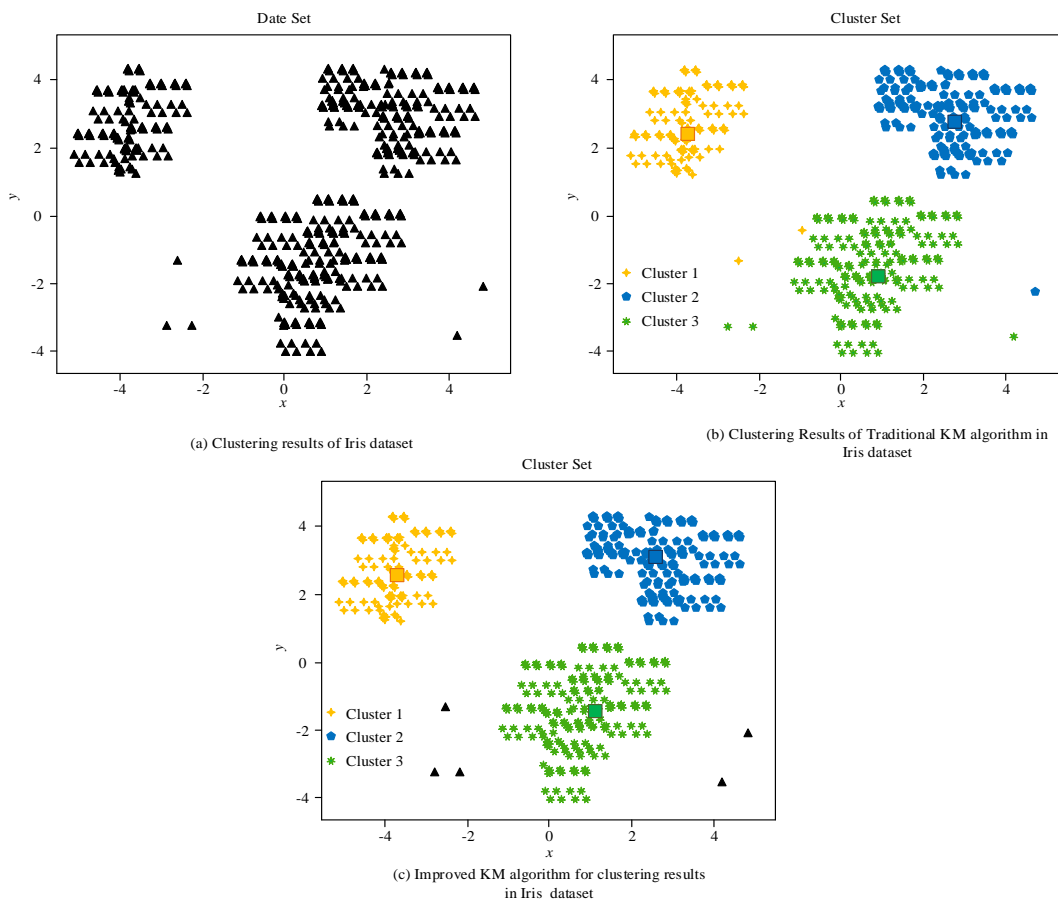


Fig. 9. The clustering results of two algorithms in the Iris dataset.

TABLE III. CLUSTER EVALUATION OF TRADITIONAL AND IMPROVED KM ALGORITHMS

Data Set	Crowd Together	Traditional KM Algorithm			Improved KM Algorithm		
		Misclassification rate (%)	Purity (%)	Clustering accuracy	Misclassification rate (%)	Purity (%)	Clustering accuracy
Wine	Cluster 1	1.2	98.8	97.9	1.3	98.7	98.2
	Cluster 2	3.5	96.5		3.5	96.5	
	Cluster 3	1.3	98.7		0	100	
	Cluster 4	2.5	97.5		2.1	97.9	
Iris	Cluster 1	2	98	97.6	0	100	100
	Cluster 2	1.1	98.9		0	100	
	Cluster 3	4	96		0	100	

### V. CONCLUSION

In the context of data ecology, the proliferation of data resources has led to the problems of information pollution and data redundancy in the supply chain service of collection resources. For solving the data redundancy issue and enhancing the collection resource supply efficiency of librarians, the study used DM technology combined with improved KM CA for designing a VCC model of library collection resource supply chain for librarians under the data ecology. The experiment illustrated that the CA of the traditional KM algorithm in the Wine dataset was 70.3% at the highest and 53.5% at the lowest, and the CA of the improved KM algorithm stable at 71%; the CA of the traditional KM

algorithm in the Iris dataset was 65% at the lowest and 89% at the highest, and the CA of the improved KM algorithm was maintained at 90%. The running time of the traditional KM algorithm in the Wine dataset was 40ms at the shortest and 115 ms at the longest, and that of the improved KM algorithm is stable at 59 ms; the running time of the traditional KM algorithm in the Iris dataset was 26 ms at the shortest and 58ms at the longest, and that of the improved KM algorithm was maintained at 53 ms. The clustering effect of the improved KM algorithm in practice was 98.2% in the Wine dataset, which was 0.3% higher than the 97.9% CA of the traditional KM algorithm, and 100% in the Iris dataset, which was 2.4% exceeding the 97.6% CA of the traditional KM algorithm. 2.4%. In summary, it illustrates that the studied

VCC model of the supply chain of library resources in view of the data ecology of the library dispensers has better resource supply efficiency, but the experiments use fewer comparison sample data. The experimental results are not objective enough, and further improvement is needed in this aspect. This article studies and explores the influencing factors and mechanisms of VCC in the supply chain of library collection resources from the perspective of data ecology and VCC. Based on the current low utilization rate of library collection resources, differentiated reader needs, and service VCC, improvement strategies are proposed. The conclusions of the study are as follows: firstly, the influencing factors of VCC in the supply chain of library resources mainly include nine major factors: resource construction and integration in the supply chain, development and application of supply chain service technology, data application of supply chain value co creation, staff literacy in the supply chain, concept recognition of supply chain VCC, construction of supply chain VCC platform, subject interaction of supply chain VCC, the reader needs of supply chain VCC and the environmental constraints of supply chain VCC. Secondly, based on the shared supply chain platform and reader decision-making procurement, a VCC model for the collection resource supply chain has been constructed. This model can not only meet the personalized needs of readers, but also improve the utilization rate of collection resources, making the value of the library more fully realized.

#### REFERENCES

- [1] Visser M, Van Eck N J, Waltman L, "Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic," *Quantitative Science Studies*, vol. 2, no. 1, pp. 20-41, 2021.
- [2] Wu X, Duan J, Pan Y & Li, M, "Medical knowledge graph: Data sources, construction, reasoning, and applications," *Big Data Mining and Analytics*, vol. 6, no. 2, pp. 201-217, 2023.
- [3] Waziri T A, Ibrahim A, "Discrete Fix Up Limit Model of a Device Unit," *Journal of Computational and Cognitive Engineering*, vol. 2, no. 2, pp. 163-167, 2022.
- [4] Cui Z, Yan C, "Deep integration of health information service system and data mining analysis technology," *Applied Mathematics and Nonlinear Sciences*, vol. 5, no. 2, pp. 443-452, 2020.
- [5] Yang Y, Jia F, Xu Z, "Towards an integrated conceptual model of supply chain learning: an extended resource-based view," *Supply Chain Management: An International Journal*, vol. 24, no. 2, pp. 189-214, 2019.
- [6] Nandi M L, Nandi S, Moya H & Kaynak, H, "Blockchain technology-enabled supply chain systems and supply chain performance: a resource-based view," *Supply Chain Management: An International Journal*, vol. 25, no. 6, pp. 841-862, 2020.
- [7] Agyabeng-Mensah Y, Ahenkorah E, Afum E, Agyemang, A. N., Agnikpe, C & Rogers, F, "Examining the influence of internal green supply chain practices, green human resource management and supply chain environmental cooperation on firm performance," *Supply Chain Management: An International Journal*, vol. 25, no. 5, pp. 585-599, 2020.
- [8] Mehrotra S, Rahimian H, Barah M, Luo, F & Schantz, K., "A model of supply-chain decisions for resource sharing with an application to ventilator allocation to combat COVID-19," *Naval Research Logistics (NRL)*, vol. 67, no. 5, pp. 303-320, 2020.
- [9] Zhao Y, Zhang C, Zhang Y, Wang, Z & Li, J, "A review of data mining technologies in building energy systems: Load prediction, pattern identification, fault detection and diagnosis," *Energy and Built Environment*, vol. 1, no. 2, pp. 149-164, 2020.
- [10] Liu J, Zhou S, "Application research of data mining technology in personal privacy protection and material data analysis," *Integrated Ferroelectrics*, vol. 216, no. 1, pp. 29-42, 2021.
- [11] Rong Z, Gang Z. An artificial intelligence data mining technology based evaluation model of education on political and ideological strategy of students. *Journal of Intelligent & Fuzzy Systems*, vol. 40, no. 2, pp. 3669-3680, 2021.
- [12] Wang J, Yang Y, Wang T, Sherratt, R. S & Zhang, J, "Big data service architecture: a survey," *Journal of Internet Technology*, vol. 21, no. 2, pp. 393-405, 2020.
- [13] Yang J, Li Y, Liu Q, Li, L., Feng, A., Wang, T & Lyu, J, "Brief introduction of medical database and data mining technology in big data era," *Journal of Evidence-Based Medicine*, vol. 13, no. 1, pp. 57-69, 2020.
- [14] Sutduean J, Singa A, Sriyakul T & Jermittiparsert, K, "Supply chain integration, enterprise resource planning, and organizational performance: The enterprise resource planning implementation approach," *Journal of Computational and Theoretical Nanoscience*, vol. 16, no. 7, pp. 2975-2981, 2019.
- [15] Lou H, "Design of college English process evaluation system based on data mining technology and internet of things," *International Journal of Data Warehousing and Mining (IJDDWM)*, vol. 16, no. 2, pp. 18-33, 2020.
- [16] Acquah I S K, Agyabeng-Mensah Y, Afum E, "Examining the link among green human resource management practices, green supply chain management practices and performance," *Benchmarking: An International Journal*, vol. 28, no. 1, pp. 267-290, 2020.
- [17] John S J, Suma P, Athira T M, "Multiset modules," *Journal of Computational and Cognitive Engineering*, vol. 1, no. 1, pp. 37-41, 2022.
- [18] Ageed Z S, Zeebaree S R M, Sadeeq M M, Kak, S. F., Yahia, H. S., Mahmood, M. R., & Ibrahim, I. M., "Comprehensive survey of big data mining approaches in cloud systems," *Qubahan Academic Journal*, vol. 1, no. 2, pp. 29-38, 2021.
- [19] Smarandache F, "Plithogeny, plithogenic set, logic, probability and statistics: a short review," *Journal of Computational and Cognitive Engineering*, vol. 1, no. 2, pp. 47-50, 2022.
- [20] Nandi S, Sarkis J, Hervani A & Helms, M. "Do blockchain and circular economy practices improve post COVID-19 supply chains? A resource-based and resource dependence perspective," *Industrial Management & Data Systems*, vol. 121, no. 7, pp. 333-363, 2021.

# A Flexible Manufacturing System based on Virtual Simulation Technology for Building Flexible Platforms

Zhangchi Sun

Engineering Training and Innovation Education Center, Shanghai Polytechnic University, Shanghai, 201209, China

**Abstract**—Flexible manufacturing systems have become relatively mature in the industrial field, representing the most advanced research achievements in the development of the manufacturing industry. But currently, there are few resources and high costs in universities to create a system that is more practical, and it cannot meet the practical teaching requirements of students in multiple majors. In response to the above issues, this study first designed a flexible manufacturing system from the overall architecture, then introduced and integrated virtual simulation technology, and utilized multi-objective genetic algorithm for cargo location optimization to improve the work efficiency of the flexible system. The research results indicate that after 213 iterations of the proposed algorithm, the iteration curve of the total objective function value tends to be stable, and the effect of cargo location optimization is relatively ideal. At this time, the total objective function value is 142.5. In addition, as the scale expands, the corresponding number of iterations for multi-objective genetic algorithm at its maximum scale is 411.2. The application effect of virtual flexible manufacturing system in practical teaching in universities is good, and visual learning methods can better attract students' attention.

**Keywords**—Flexible platform; virtual simulation technology; manufacturing control system; multi-objective genetic algorithm; slotting optimization

## I. INTRODUCTION

With the advent of the Fourth Industrial Revolution and the rapid development of the social economy, people have put forward higher requirements for the richness of the industrial industry. This has led to a shortened lifespan of industrial products, and currently traditional single production and manufacturing technologies are no longer able to meet industrial production needs [1-2]. Therefore, to improve industrial production efficiency and meet the flexible requirements of production manufacturing systems, Flexible Manufacturing Systems (FMS) have emerged and received widespread support in related fields. FMS is a set of CNC machine tools and other automated process equipment, which is an organic combination of computer information control system and automatic material storage and transportation system. It has the advantages of high equipment utilization rate, high product quality, flexible operation, stable production capacity, and large product response capacity [3-4]. However, universities lack relevant technology and funding to learn flexible manufacturing technology, resulting in a lack of theoretical knowledge and practical application for students in related majors [5-6]. And the cost of the flexible

manufacturing system close to the actual production is very expensive, which cannot meet a large number of teaching tasks, which hinders the path of cultivating mechanical automation talents in Colleges and universities, and does not meet the needs of social development. To reduce damage to flexible manufacturing equipment caused by improper operation by students and avoid personal injury, this study introduces Virtual Simulation Technology (VST) to construct a Virtual Flexible Manufacturing (VFM) system, to meet the combination of physical control system of virtual equipment and practical teaching. This paper aims to address the limitations of flexible physical systems in practical training and teaching in universities, create FMS that is more in line with practical production, and cultivate the professional and practical abilities of comprehensive talents. There are two main innovative points in the study. The first point is to propose a VFM system that integrates VST, which can interact with physical control systems, receive control commands, and feedback virtual sensor signals. The second point is to use multi-objective genetic algorithm (MOG) for cargo location optimization to improve the efficiency of flexible systems. The structure of the study mainly consists of four parts. The first part is a review of relevant research results. The second is about the design of VFM that integrates VST, as well as the optimization of cargo space in the VFM system based on MOG. The third is to verify the effectiveness and applicability of the proposed method; the final part is a summary of the entire content.

## II. RELATED WORK

With the acceleration of informatization and industrialization, the performance of industrial products is becoming increasingly complex, and traditional physical manufacturing systems are unable to meet the current production needs of industrial products. FMS has emerged as the times require. It enhances the competitiveness of enterprises, which also puts higher requirements on the application abilities of relevant practitioners and professional students. Numerous scholars have conducted in-depth research and exploration on this topic. H. Wang et al. found that there are many factors unrelated to processing in flexible manufacturing systems, resulting in significant differences in the formulation and implementation of production plans. Therefore, a study proposed an improved genetic algorithm with local search to optimize scheduling data based on discrete manufacturing enterprises. The experimental results show that compared with the current scheduling strategy of

the enterprise, the scheduling strategy proposed by this algorithm has an average improvement of 29.61% in minimizing completion time, 44.8% in minimizing transportation time, and 44.64% in machine load balancing [7]. C. Zhu et al. proposed an integrated optimization method to plan and schedule a hybrid flexible manufacturing system for producing nylon components, in order to minimize energy consumption and completion time. The effectiveness and feasibility of the proposed method were verified through practical cases [8]. Setiawan A proposes an FMS based on a stacker crane model to establish a learning production system. The maximum error of this model in the x-axis and y-axis directions is 2mm [9]. Daniyan I et al. proposed an FMS including assembly, Lean manufacturing, logistics and quality assurance to adapt to the dynamic of manufacturing operations. The system can properly perform the sequence of assembly and quality assurance operations with minimal interruption and manual intervention during the manufacturing of rail car components [10].

VST is the mapping activity of a virtual simulation environment on a computer. By constructing actual device models in virtual environments and utilizing the excellent data processing capabilities of computers, VST can achieve the same results as real application scenarios, making it a new technology in various fields. M. Wei et al. innovatively introduced a virtual simulation platform to teach the power system analysis course in order to improve students' understanding of knowledge, hands-on ability, and research skills. After conducting the reform teaching, students were very satisfied with the new teaching method and believed that it could effectively stimulate students' enthusiasm for learning [11]. J. Wood et al. found that the confidence and stress levels of providers are related to the survival outcomes of patients. Therefore, they innovatively utilized virtual simulation technology to create training opportunities to enhance providers' confidence and reduce stress. The research results showed that this method can effectively enhance trust in resuscitation training and reduce stress [12]. Wang Y and others proposed a multi Kinect fusion algorithm to achieve robust full body tracking in virtual reality assisted assembly simulation, and applied distributed computing to improve computational efficiency in the algorithm. Compared with other similar algorithms, this algorithm has better fusion performance [13]. To improve the performance of online art Design education system, Yang C built an online art Design education system based on 3D VST. The constructed model can meet the current needs of online art education, and the functional modules can also be continuously optimized in the future [14].

In summary, there are many research results on FMS and VST, but currently there is relatively little research on VST for FMS in universities. To shorten the design cycle of flexible production lines and improve their design accuracy, this study introduced VST to construct a VFM system.

### III. DESIGN OF VFM INTEGRATING VST

At present, China's research and application on FMS is rapidly advancing. But there is no flexible manufacturing platform in various universities in China that meets the actual

production line for teaching, resulting in the inability of students to apply their professional knowledge in practice. Therefore, the research first designs the overall scheme of VFM system, then uses VST technology to build the system, and finally proposes a MOG algorithm for VFM system to optimize the storage location.

#### A. Design of FMS Integrated with VST Technology

In order to design the VFM system for teaching and training links and scientific research projects, the research requires that it cannot only carry out single electrical course training, but also have the ability to integrate multiple courses. Based on the teaching tasks of college students, the following requirements are proposed for the VFM system. Firstly, it needs to not only have flexibility in processing and manufacturing, but also consider flexibility in teaching practice. Secondly, in the teaching curriculum, the theoretical content includes various mechanical and software control methods, so it is necessary to achieve teaching and practical training for different purposes in the design of VFM systems. Finally, flexible manufacturing technology is also constantly developing, and the speed of system updates and iterations is also accelerating. This requires the design of the VFM system to meet the requirements of photography technology updates to achieve the upgrade of the system. In university teaching, the physical manufacturing teaching system may have drawbacks, such as high teaching costs, resource shortages, slow security and system upgrade iterations in student practice [15-16]. To address the above issues, the study introduces VST in the design of VFM, enabling students to conduct virtual simulations of VFM through virtual simulation software. In addition, the research selected the current mainstream computer 3D display technology and placed virtual devices with the same functions and attributes as flexible devices in a virtual environment to build a VFM system. At the same time, by adding virtual devices to upgrade and optimize the system, each computer can be seen as a set of flexible systems, which effectively solves the problems of high cost and limited resources in traditional flexible systems. According to its functions, it can be divided into three parts: logistics warehousing, production manufacturing, and control systems. The logistics warehousing system in the VFM system is its advantage, with high intelligence and freedom, which can distinguish various types of workpieces for mixed transportation. This link can use the information processing of computers and controllers, and perform intelligent control of multiple devices at the same time; The system can also transport processed parts in different functional areas, and the three-dimensional warehouse can improve its storage space utilization and workpiece processing efficiency through an intelligent automatic warehouse information relationship system. The production and manufacturing system includes flexible production lines, six axis CNC machining centers, robots, and inspection devices. The control system is the command core of VFM. Design VFM by analyzing its impact on the functionality and work efficiency of the system, and study the optimization of warehouse locations in the system to improve the efficiency of flexible system work by improving the efficiency of goods in and out. From this, the VFM design process in Fig. 1 can be obtained.

The virtual development platform used in this study is a specialized simulation software for electromechanical systems. It is based on the latest virtual debugging and simulation technology and provides users with a fully open virtual device development platform. The 3D model is based on the development of mainstream 3D modeling software, and uses the latest 3D rendering technology and Physics engine to show the most realistic simulation effect. Compared with other simulation software, electromechanical integration simulation software and physical controllers have better information exchange capabilities, which can simplify the implementation of external controllers' motion control of virtual devices in virtual software. The hardware equipment in the designed VFM system uses computers and HTC VIVE. Computers provide a running environment for simulation software; HTC VIVE is the implementation of Virtual Reality (VR) functionality in simulation systems. From this, the overall framework of the virtual simulation system can be obtained, as shown in Fig. 2, which is composed of a model layer, a data layer, and a human-machine interaction layer.

### B. Construction of FMS Integrating VST

The proposed VFM system has novelty. Virtual flexible devices not only have the same mechanical structure and

motion mode as actual devices, but also interact with physical control systems for signals. It also has the function of receiving control commands and feedback virtual sensor signals. In the 3D modeling process of the VFM system, the modeling results generate a large amount of redundancy, and exporting them can also generate a large amount of data. After being imported into simulation software, it will occupy a large amount of computer memory, causing problems such as crying and incomplete animation display when the virtual simulation system runs. Therefore, this study will achieve lightweight processing by converting 3D models into 3DXML format. Therefore, the research realizes the lightweight processing by converting the 3D model into 3DXML format, which only contains the entity information of the 3D model and the assembly features of the mechanical structure, and its storage space is reduced by 90% compared with the traditional CAD and manifest formats. There are many 3D shapes in 3D modeling, which are not necessary for the simulator and can greatly reduce the display effect. Therefore, the most commonly used method for simplifying 3D models is the edge folding algorithm. Its principle is as follows: let a plane  $Q$  be (1) [17].

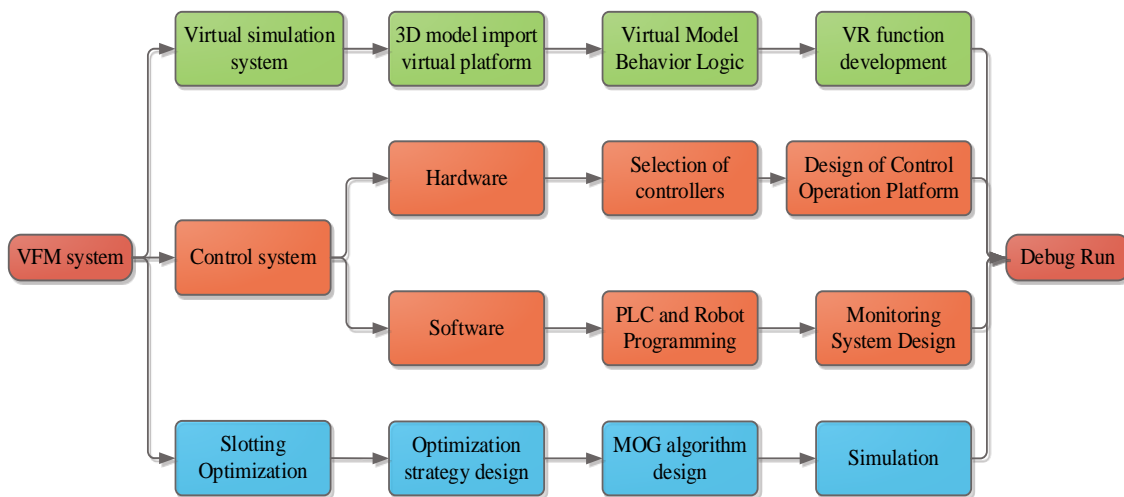


Fig. 1. VFM system design process.

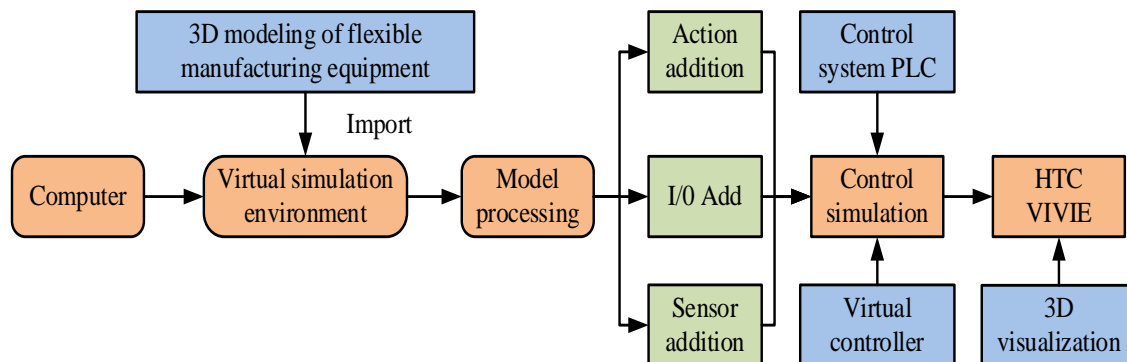


Fig. 2. The overall framework of virtual simulation systems.

$$\begin{cases} ax+by+cz+d=0 \\ a^2+b^2+c^2=1 \end{cases} \quad (1)$$

In (1),  $d$  represents a constant. By setting the coordinate of point  $p$  as  $(x, y, z)$ , the distance from it to  $Q$  can be obtained, as (2).

$$d(p) = |ax+by+cz+d| \quad (2)$$

By (2),  $Q = [abcd]^T$  can be gaibed. If the other vertex is  $\bar{g} = [xyz1]^T$ , the square of the distance from  $\bar{g}$  to  $Q$  is obtained as (3).

$$d_Q^2(\bar{g}) = \bar{g}^T D_Q \bar{g} \quad (3)$$

In (3),  $D_Q$  is a Symmetric matrix of  $4 \times 4$ , as (4).

$$D_Q = \begin{bmatrix} a^2 & ab & ac & ad \\ ab & b^2 & bc & bd \\ ac & bc & c^2 & cd \\ ad & bd & cd & d^2 \end{bmatrix} \quad (4)$$

When folding edge  $(g_1, g_2)$  to  $\bar{g}$ , the quadratic error matrix of  $\bar{g}$  is obtained. Then simplify the operation and take matrix  $L(g_1)+L(g_2)$  as the quadratic error matrix between the new vertex and  $(g_1, g_2)$  to obtain the folding cost of  $(g_1, g_2)$ , as (5).

$$\Delta(\bar{g}) = \bar{g}^T (L(g_1 + Lg_2)) \bar{g} \quad (5)$$

To perform edge folding operation, a position must be selected for  $\bar{g}$ . Since the Error function is a Quadratic function, the minimum value of  $\Delta(\bar{g})$  can be calculated from its partial derivative, as (6).

$$\frac{\partial \Delta(\bar{g})}{\partial x} = \frac{\partial \Delta(\bar{g})}{\partial y} = \frac{\partial \Delta(\bar{g})}{\partial z} = 0 \quad (6)$$

Assuming matrix  $h$  is (7).

$$h = \begin{bmatrix} h_{11} & h_{12} & h_{13} & h_{14} \\ h_{12} & h_{22} & h_{23} & h_{24} \\ h_{13} & h_{23} & h_{33} & h_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (7)$$

If  $h$  is an invertible matrix, the new vertex  $\bar{g}'$  can be obtained as (8).

$$h = \begin{bmatrix} h_{11} & h_{12} & h_{13} & h_{14} \\ h_{12} & h_{22} & h_{23} & h_{24} \\ h_{13} & h_{23} & h_{33} & h_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \quad (8)$$

If  $h$  is an irreversible matrix, select two breakpoints or key points of  $(g_1, g_2)$  as new vertices. The 3D models of different flexible devices are processed and rendered in a lightweight manner to obtain a virtual model, as Fig. 3.

Fig. 3 includes a three-dimensional warehouse, CNC machining center, conveyor belt, conveyor robot, and visual gripper robot. The development of VR functions in virtual simulation systems requires the application of HTC VIVE virtual reality devices for assistance. Compared with the traditional VR equipment, students can freely carry out mechanical equipment through the Head-mounted display and can observe the mechanical structure closely. The data collection and transmission structure of HTC VIVE is Fig. 4. The connection between HTC VIVE hardware and computer ports is the foundation for VR interaction.

### C. Optimization of Cargo Location for VFM

To further explore methods to improve the efficiency of VFM work, this study focuses on the problem of low efficiency in managing goods in flexible systems. The VFM system is used as the research object and the MOG algorithm is used for cargo location optimization. After running the VFM system for a period of time, due to the use of automatic mode, manual mode, and cargo in and out functions during the teaching process, the flexible system may experience scattered storage of goods and unreasonable distribution of cargo locations. Therefore, the study introduces MOG to optimize the allocation of warehouse locations in the VFM system, thereby improving the warehousing efficiency of the warehouse and ultimately achieving the goal of improving the efficiency of VFM work. The research will optimize the efficiency of goods in and out of storage, as well as the stability of cargo locations. The former increases the frequency of goods entering and leaving the warehouse in the warehousing system, and allocates them to a location closer to the warehouse entrance and exit, in order to shorten the time for goods entering and leaving the warehouse. In the optimization of cargo space stability, in order to better align with practical applications, it is necessary to study adding physical attributes such as mass and friction to the virtual cargo model. Therefore, the study also needs to consider the stability of warehouse shelves in different cargo weights. In response to the above, lighter goods will be prioritized in the upper part of the shelf, while heavier goods will be allocated in the lower part of the shelf. This placement method not only improves the stability of the shelf, but also reduces the power consumption of the warehouse during inbound and outbound operations. Genetic algorithm (GA) currently performs well in global search and can effectively reduce the probability of getting stuck in local search for the optimal solution. Compared with other algorithms, this algorithm is more suitable for solving complex combinatorial problems and can search for the optimal solution more quickly. Due to the multi-objective nature of the cargo location optimization problem in VFM, the research chose the MOG algorithm based on GA for solution [18-20]. Fig. 5 shows the structural flow of a standard GA.

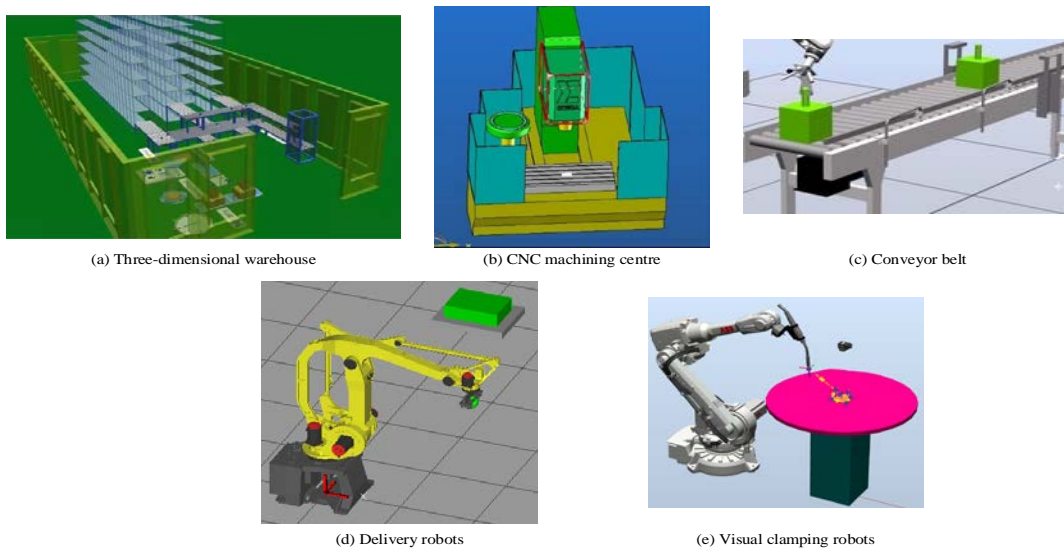


Fig. 3. Schematic diagram of virtual lean meat equipment.

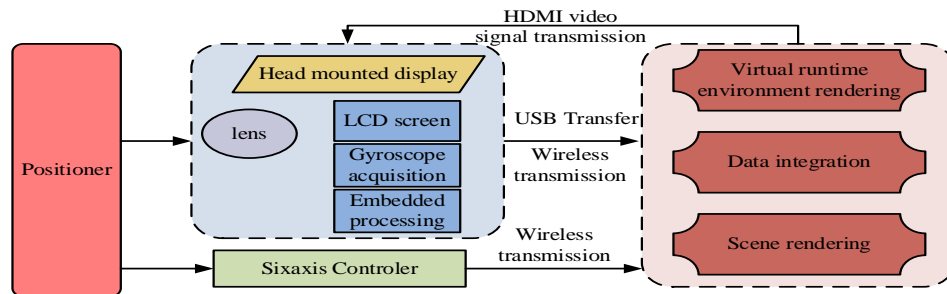


Fig. 4. Data Collection and transmission structure of HTC VIVE.

Before optimizing the storage space, it is first necessary to determine the optimization object and its key information attributes. According to the three-dimensional warehouse of the VFM system, the objects of cargo location optimization are the stacker, the goods processed by the system, and the cargo spaces included in the shelves. Key information includes the X-axis and Y-axis speeds of the stacker, as well as the origin position. The goods include weight, model, and delivery rate; The storage location contains a number. The study designed three types of processed goods in VFM and added physical properties to them, so that the goods had different qualities and met optimization requirements, as Fig. 6.

The objective of cargo location optimization is to optimize the efficiency of goods in and out of storage and the stability of cargo locations. The multi-objective mathematical model established is (9).

$$\begin{cases} f = \omega_1 f_1 + \omega_2 f_2 \\ f_1 = \sum_{x=1}^a \sum_{y=1}^b \sum_{z=1}^c R_{xyz} t_{xyz} \\ f_2 = \sum_{x=1}^a \sum_{y=1}^b \sum_{z=1}^c (m_{xyz} \times n_{xyz} \times z) \\ \omega_1 + \omega_2 = 1 \end{cases} \quad (9)$$

In (9),  $f_1$  and  $f_2$  are the objective function values for optimizing the efficiency of goods entering and exiting the warehouse and optimizing the stability of the storage space, respectively.  $\omega_1$  and  $\omega_2$  are the corresponding weights for optimizing the efficiency of goods in and out of storage and optimizing the stability of storage locations.  $R_{xyz}$  represents the shipment rate of the goods at location  $(x, y, z)$ .  $t_{xyz}$  represents the time it takes for the goods to move from the storage location to the warehouse.  $m_{xyz}$  and  $n_{xyz}$  correspond to the weight and quantity of the goods. The steps to optimize using the MOG algorithm are as follows: first, use integers for encoding, so that the integers composed of  $(x, y, z)$  correspond to the rows, columns, and layers of the shelves. Secondly, the fitness function is determined. According to the multi-objective mathematical model, the objective function is positive and the minimum demand solution is obtained. The calculation is (10).

$$\begin{cases} Fit(x, y, z) = \sum_{i=1}^n \omega_i \times \frac{1}{f_i - f_{\min} + 1} \\ \begin{cases} 1 \leq x \leq a \\ 1 \leq y \leq b \\ 1 \leq z \leq c \end{cases} \end{cases} \quad (10)$$



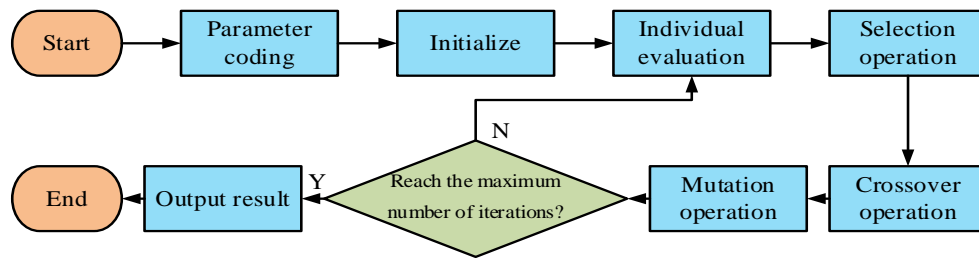


Fig. 5. The process of standard GA algorithm.

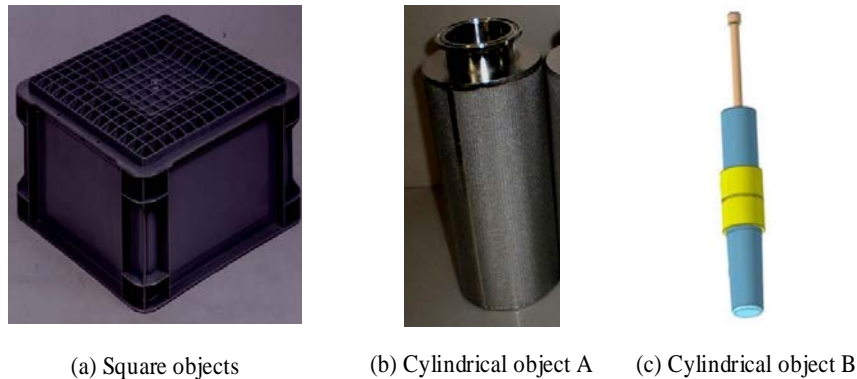


Fig. 6. Three virtual cargo schematic diagrams for VFM system design.

Finally, the selection operator and other related parameters are determined, and the selection operator uses the roulette wheel selection method.

#### IV. RESULT ANALYSIS OF VFM FUSED WITH VST

To verify the performance of the VFM system and its effectiveness in practical teaching applications, this chapter first analyzes the performance of the system based on the MOG algorithm and its effectiveness in cargo location optimization, and then analyzes the application of VFM in practical teaching.

##### A. Analysis of Cargo Location Optimization Results based on MOG Algorithm

Firstly, to verify the performance of cargo location optimization based on MOG, MATLAB software is used for simulation experiments. The research will evaluate the optimization results based on the objective function value and cargo distribution. In addition, in order to verify the performance of the algorithm more scientifically, Multi objective optimization algorithm (MOO) commonly used at present is selected for comparative experiments, such as Multi objective Particle Swarm (MOPS), Multi objective Evolutionary algorithm (MOE) and Parallel Single Ended Inheritance (PSEI). Table I shows the cargo attributes and parameter settings of MOG.

From Fig. 7, it can be observed that after 213 iterations of the proposed MOG algorithm, the iteration curve of the total objective function value tends to stabilize, and the effect of cargo location optimization is relatively ideal. At this time, the corresponding total objective function value is 142.5. MOPS

needs to iterate 307 times before the change curve starts to stabilize, with a stable function value of 148.3; MOE stabilized after 256 cycles, with a value of 152.7; PSEI requires 286 iterations to reach a stable state, with a stable value of 150.2. The above results show that the MOG algorithm can obtain lower total objective function values in smaller iterations.

The paper uses MATLAB software for cargo location optimization operations using different multi-objective optimization algorithms.

To further explore the optimization effects of different MOOs in different scale problems, four optimization problems with different scales were set up, as displayed in Table II. From it, the MOG algorithm has the best optimization performance at different scales, followed by MOE and MOPE. And as the target scale problem expands, the optimization performance of each algorithm also shows varying degrees of decline. This is because as the scale expands, the objective function value of the original plan shows an exponential increase, so the decrease in optimization effect is within a reasonable range. From the perspective of convergence, when the scale is small, the iteration times of the four algorithms are not significantly different, because the solution to the optimal solution is relatively simple at small scales. However, the expansion of scale has demonstrated the advantage of MOG in iteration speed. At the maximum scale, the corresponding number of iterations is 411.2, which improves the efficiency of MOPS, MOE, and PSEI by 17.3%, 36.7%, and 40.9%. In summary, the MOG algorithm can effectively improve the probability of local optimization and accelerate convergence, with good cargo location optimization performance.

TABLE I. RELEVANT PARAMETER SETTINGS AND CARGO ATTRIBUTE INFORMATION OF MOG ALGORITHM

Setting of relevant parameters for MOG algorithm	Parameter	Set value	Parameter	Set value
	Number of shelf rows	3	Population size	300
	Number of shelf columns	8	Maximum Number Of Iterations	500
	Number of shelves	6	Crossover probability	0.8
	Location length/m	0.5	Mutation probability	0.08
	Stacker X-axis direction speed/(m/s)	0.6	Shelf spacing/m	0.9
	Stacker Y-axis direction speed/(m/s)	0.5	$\omega_1$	0.5
	Stacker Z-axis direction speed/(m/s)	0.3	$\omega_2$	0.5
Property information of goods	Type of goods	Cargo mass/kg	Delivery rate of goods	
	Column cargo E	0.8		0.5
	Column cargo H	0.6		0.3
	Column cargo I	0.4		0.1

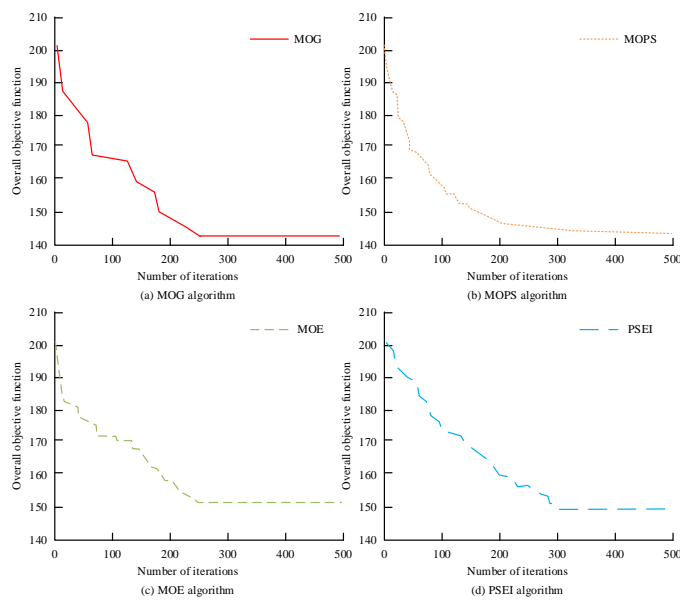


Fig. 7. The variation curve of the total objective function value for cargo location optimization using different MOO.

TABLE II. OPTIMIZATION RESULTS OF VARIOUS MOO IN PROBLEMS OF DIFFERENT SCALES

Number of goods/number of goods		25/120	50/203	75/547	100/810
MOG	$f_1$	99.50	91.34	108.98	132.49
	$f_2$	189.2	193.7	185.3	178.5
	Iterations	194.2	2993.5	336.6	411.2
MOPS	$f_1$	99.59	93.43	112.42	147.43
	$f_2$	189.3	199.0	187.3	234.2
	Iterations	218.8	323.5	419.6	482.3
MOE	$f_1$	102.3	93.6	124.7	139.2
	$f_2$	152.4	198.2	252.8	290.5
	Iterations	276.1	421.5	488.2	562.2
PSEI	$f_1$	112.33	85.43	120.32	141.08
	$f_2$	264.1	223.4	267.8	287.9
	Iterations	284.5	348.3	385.8	579.2

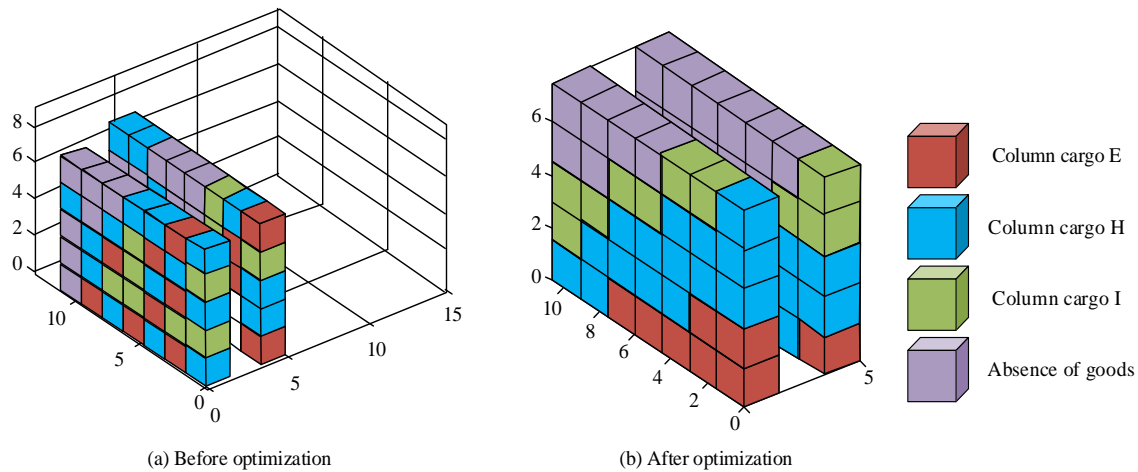


Fig. 8. Effect diagram of the distribution of warehouse locations before and after optimization based on MOG.

To verify the effectiveness of MOG in optimizing warehouse locations, Fig. 8 was developed. Fig. 8(a) and 8(b) represent the distribution of goods before and after the optimization of the cargo location. In Fig. 8(a), the goods were placed in a chaotic and disorderly manner in the warehouse of the VFM system, and there were no goods placed at the bottom of the shelves. This method of placing goods results in poor storage stability and has a negative impact on the efficiency of Tiger House's inbound and outbound operations. In Fig. 8(b), after MOG optimization, some goods are placed in a regular and uniform manner in the warehouse, with heavy items placed at the bottom of the storage space and light items placed at the top of the storage space. The results show that MOG based cargo location optimization can ensure both the efficiency of goods in and out of storage and the stability of shelves.

### B. Application Analysis of VFM

To explore the application of VFM in practical training teaching, the study first prepares three hardware devices: computer, HTC VIVE, and control console. The hardware connection of this system includes the communication connection between the console and the computer, the use of TCP/IP protocol and network cables to complete the communication connection between the computer and PLC, and the connection between the computer and HTC VIVE. Connect the VR helmet to the computer graphics card and

transfer data through HDMI and USB cables. After completing the hardware connection, communication connections between different software on the platform can be made. Simply add PLC geology to complete the communication connection.

During the operation of the VFM system, a schematic diagram of the main virtual flexible equipment can be obtained, as Fig. 9. The main flexible equipment working process includes the three-dimensional warehouse outbound, the gripping operation of the conveying robot, and the machining process of the CNC machine tool. After the construction of HTC VIVE is completed, the VFM system can be visualized in 3D.

To analyze the application of VFM in various universities, a corresponding usability and satisfaction evaluation table was set up, and a survey was conducted on 5000 students using the system in universities across the country. The results are listed in Fig. 10. Most students have a good evaluation of the use of VFM, with scores exceeding 18 for different evaluation questions, and a total score of  $93.34 \pm 3.09$ . This indicates that the user experience of the system is excellent. Students believe that the use of this system can enhance their learning confidence, and most students are willing to continue using the system for learning. The results show that the VFM system is well applied in universities and is widely loved by students.

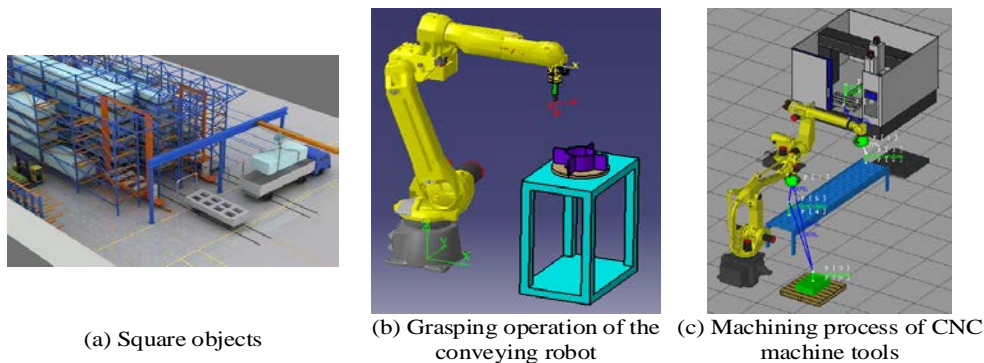


Fig. 9. Schematic diagram of virtual flexible equipment operation.

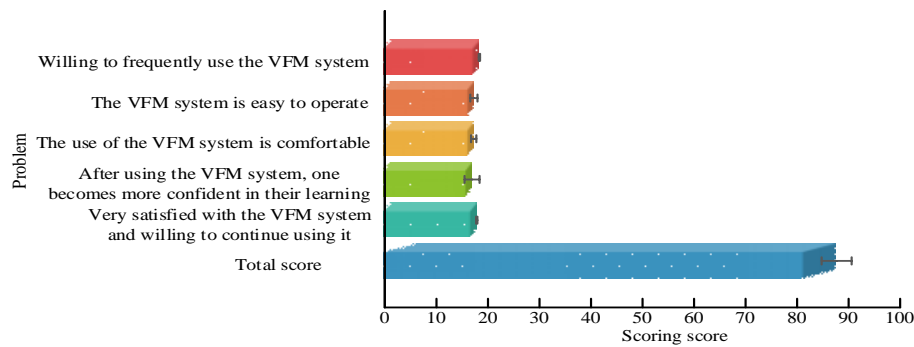


Fig. 10. Availability and satisfaction analysis of VFM system.

## V. CONCLUSION

Many universities in China lack the application of FMS that is compatible with actual production lines in practical training and teaching, and due to issues such as resource shortages and high documentation costs, actual FMS cannot meet teaching tasks. To address the issue of the inability of university students to translate their knowledge into practical applications, the study first completed the overall design of the VFM system, followed by the design of the VST-VFM system, and finally optimized the efficiency of cargo allocation. Experimental results show that compared with other mainstream algorithms, the proposed MOG algorithm has the least number of iterations of 213. At this time, the iterative change curve of the total objective function value tends to be stable, and the effect of cargo location optimization is ideal. The average value of the corresponding total objective function is 142.5, which shows that the proposed method can not only improve the efficiency of goods in and out of the warehouse, but also enhance the stability of the shelf. With the expansion of the scale, the number of iterations of the proposed MOG algorithm is 411.2 at the maximum scale, which is 17.3%, 36.7% and 40.9% higher than that of mops algorithm, MOE algorithm and PSEI algorithm. The placement of goods in the warehouse optimized by MOG algorithm is regular and uniform, which means that the proposed method can effectively improve the probability of local optimization, accelerate convergence, and have better performance of cargo location optimization. In the evaluation results of the availability and satisfaction of VFM system applied in Colleges and universities, the total score is  $93.34 \pm 3.09$ , which indicates that students are very satisfied with the experience of the system and believe that it can effectively improve their learning confidence. In summary, the VFM system proposed in the study has excellent performance and has good application effects in practical training and teaching in universities. However, there are still shortcomings in the research. The constructed VFM only has basic functions for actual processing and production. In future research, VST can be used to introduce other equipment or models for improvement.

## REFERENCES

- [1] G. Zang, P. Sun, A. Elgowainy, and M. Wang, "Technoeconomic and Life Cycle Analysis of Synthetic Methanol Production from Hydrogen and Industrial Byproduct CO<sub>2</sub>," *Environmental Science and Technology*, vol. 55, no. 8, pp. 5248-5257, 2021.
- [2] S. Ruidas, M. R. Seikh, and P. K. Nayak, "A production inventory model for high-tech products involving two production runs and a product variation," *Journal of Industrial and Management Optimization*, vol. 19, no. 3, pp. 2178-2205.
- [3] M. N. A. Khalid, and U. K. Yusof, "An Improved Immune Algorithms for Solving Flexible Manufacturing System Distributed Production Scheduling Problem Subjects to Machine Maintenance," *International Journal of Mathematical Models and Methods in Applied Sciences*, vol. 15, pp. 17-25, 2021.
- [4] P. Dubey, P. Chauhan, P. Shekhar, P. Agarwal, and P. Sharma, "Quantisation of driving enablers enabling transition from traditional to flexible manufacturing system," *International Journal of Manufacturing Technology and Management*, vol. 35, no. 2, pp. 89-108, 2021.
- [5] Z. P. Yin, Y. A. Huang, H. Yang, J. Chen, Y. Duan, and W. Chen, "Flexible electronics manufacturing technology and equipment," *Science China Technological Sciences*, vol. 65, no. 9, pp. 1940-1956, 2022.
- [6] X. Y. Wang, Q. Zhao, B. G. Hyun, L. Yu, L. Ma, H. Liu, J. Wang, and C. B. Park, "Flexible Poly(ether-block-amide)/Carbon Nanotube Composites for Electromagnetic Interference Shielding," *ACS Applied Nano Materials*, vol. 5, no. 5, pp. 7598-7608, 2022.
- [7] H. Wang, Y. Wang, X. Lv, C. Yu, and H. Jin, "Genetic Algorithm with Local Search for the Multi-Target Scheduling in Flexible Manufacturing System," *Journal of circuits, systems and computers*, vol. 31, no. 16, pp. 123-149, 2022.
- [8] C. Zhu, G. Sun, and K. Yang, "(12 8213 PROOFREAD (NG))A COMPREHENSIVE STUDY ON INTEGRATED OPTIMIZATION OF FLEXIBLE MANUFACTURING SYSTEM LAYOUT AND SCHEDULING FOR NYLON COMPONENTS PRODUCTION," *International Journal of Industrial Engineering*, vol. 29, no. 6, pp. 979-1001, 2022.
- [9] A. Setiawan, "Perancangan Model Stacker Crane Flexible Manufacturing System untuk Pembelajaran di Institusi Pendidikan," *Jurnal Rekayasa Sistem Industri*, vol. 10, no. 1, pp. 45-54, 2021.
- [10] I. Daniyan, K. Mpofo, B. Ramatsetse, E. Zeferino, and E. Sekano, "Design and simulation of a flexible manufacturing system for manufacturing operations of railcar subassemblies," *Procedia Manufacturing*, vol. 54, no. 2, pp. 112-117, 2021.
- [11] M. Wei, H. Zhang, and T. Fang, "Enhancing the course teaching of power system analysis with virtual simulation platform:," *International Journal of Electrical Engineering & Education*, vol. 60, no. 3, pp. 289-312, 2023.
- [12] J. Wood, L. Ebert, J. Duff, "Implementation Methods of Virtual Reality Simulation and the Impact on Confidence and Stress When Learning Patient Resuscitation: An Integrative Review" *Clinical Simulation in Nursing*, vol. 66, pp. 5-17, 2022.
- [13] Y. Wang, F. Chang, Y. Wu, Z. Hu, L. Li, P. Li, P. Lang, and S. Yao, "Multi-Kinects fusion for full-body tracking in virtual reality-aided assembly simulation," *International Journal of Distributed Sensor Networks*, vol. 18, no. 5, pp. 625-636, 2022.
- [14] C. Yang, "Online Art Design Education System Based On 3D Virtual Simulation Technology," *Journal of Internet Technology*, vol. 22, no. 6, pp. 1419-1428, 2021.

- [15] B. Arthaya and V. Ivan, "Preliminary Design of 3D Printing Prosthetic Hand," *Journal of Advanced Manufacturing Systems*, vol. 22, no. 1, pp. 67-84, 2023.
- [16] M. A. Nanfeng, X. F. Yao, and K. S. Wang, "Current status and prospect of future internet-oriented wisdom manufacturing," *SCIENTIA SINICA Technologica*, vol. 52, no. 1, pp. 55-75, 2022.
- [17] S. Gogos, A. Touzell, and L. B. Lerner, "What we know about intra-operative radiation exposure and hazards to operating theatre staff: A systematic review," *ANZ Journal of Surgery*, vol. 92, no. 1-2, pp. 51-56, 2022.
- [18] S. Nimrah and S. Saifullah, "Context-Free Word Importance Scores for Attacking Neural Networks," *Journal of Computational and Cognitive Engineering*, vol. 1, no. 4, pp. 187-192, 2022.
- [19] J. Zan, "Research on robot path perception and optimization technology based on whale optimization algorithm," *Journal of Computational and Cognitive Engineering*, vol. 1, no. 4, pp. 201-208, 2022.
- [20] M. Barma and U. M. Modibbo, "Multiobjective mathematical optimization model for municipal solid waste management with economic analysis of reuse/recycling recovered waste materials," *Journal of Computational and Cognitive Engineering*, vol. 1, no. 3, pp. 122-137, 2022.

# Enhanced Plagiarism Detection Through Advanced Natural Language Processing and E-BERT Framework of the Smith-Waterman Algorithm

Dr. Franciskus Antonius<sup>1\*</sup>, Myagmarsuren Orosoo<sup>2</sup>, Dr. Aanandha Saravanan K<sup>3</sup>, Dr. Indrajit Patra<sup>4</sup>, Dr. Prema S<sup>5</sup>  
Lecturer at School of Business and Information Technology STMIK LIKMI, Bandung Indonesia<sup>1\*</sup>  
School of Humanities and Social Sciences-Mongolian National University of Education, Mongolia<sup>2</sup>  
Department of ECE-Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology<sup>3</sup>  
An Independent Researcher, PhD from NIT Durgapur, West Bengal, India<sup>4</sup>  
Assistant Professor, Dept. of English-Panimalar Engineering College, Poonamalle, Chennai, India<sup>5</sup>

**Abstract**—Effective detection has been extremely difficult due to plagiarism's pervasiveness throughout a variety of fields, including academia and research. Increasingly complex plagiarism detection strategies are being used by people, making traditional approaches ineffective. The assessment of plagiarism involves a comprehensive examination encompassing syntactic, lexical, semantic, and structural facets. In contrast to traditional string-matching techniques, this investigation adopts a sophisticated Natural Language Processing (NLP) framework. The preprocessing phase entails a series of intricate steps ultimately refining the raw text data. The crux of this methodology lies in the integration of two distinct metrics within the Encoder Representation from Transformers (E-BERT) approach, effectively facilitating a granular exploration of textual similarity. Within the realm of NLP, the amalgamation of Deep and Shallow approaches serves as a lens to delve into the intricate nuances of the text, uncovering underlying layers of meaning. The discerning outcomes of this research unveil the remarkable proficiency of Deep NLP in promptly identifying substantial revisions. Integral to this innovation is the novel utilization of the Waterman algorithm and an English-Spanish dictionary, which contribute to the selection of optimal attributes. Comparative evaluations against alternative models employing distinct encoding methodologies, along with logistic regression as a classifier underscore the potency of the proposed implementation. The culmination of extensive experimentation substantiates the system's prowess, boasting an impressive 99.5% accuracy rate in extracting instances of plagiarism. This research serves as a pivotal advancement in the domain of plagiarism detection, ushering in effective and sophisticated methods to combat the growing spectre of unoriginal content.

**Keywords**—Natural language processing; encoder representation from transformers; document to vector + logistic regression

## I. INTRODUCTION

When someone exhibits another individual's software code like their own, whether purposefully or accidentally, while giving them due credit, this is known as plagiarized [1]. Plagiarism is an act of appropriating another individual's original using one's language and thoughts is seen as a breach of morality [2]. "The process or procedure of creating a different person's piece and thought, and presenting as one's

own; artistic thievery" is the meaning of unoriginality in the sense of lexicon. The act of duplicating existing music that is protected by copyright unauthorized authorization is known as music copyright infringement, and it is a hotly contested issue. In certain circumstances, the significant quantity of money at risk elevates the significance of the scenario [3]. Given the speed at which information can be shared via global platforms for collaborative engagement, writers have been motivated to conduct the chosen method of research over the Internet. Plagiarizing ideas from other individuals or research without giving due credit, plagiarism has had a negative impact. With a focus on text mining, NLP, academic literature norms, as well as several unresolved problems with standards and borderline sets, finding plagiarism is currently one of the most crucial occupations [4] [5] [6]. These foundational approaches possess great promise for addressing a variety of NLP issues, such as natural language understanding (NLU) and natural language generation (NLG), as well as potentially creating the foundation for artificial general intelligence (AGI) [7] [8]. Syntax-based and semantic-based plagiarism detection methods are the two categories into which they fall. Exemplary syntax-based methods include string comparison, AST (Abstract Syntax Tree) comparison, and token comparison. Illustrations of semantic-based methods include PDG (Programme Dependence Graph) comparing [9].

Numerous advantages include the large amount of information available on the internet in a variety of languages, as well as the accessibility of tools like engines for searching and knowledge bases, but copying has also grown. Plagiarism is the use of another investigator's ideas, substance, or results without their permission and its attribution to oneself [10]. This denies the initial investigator access to the findings of his study and makes it challenging to hunt down content, concepts, and arguments [11]. Cross-language copying is one kind of plagiarism, and it has become more prevalent as the technology for translation has advanced. To solve this issue, automated cross-language recognition of plagiarism technologies is crucial [12]. The problem of plagiarism in educational environments is not new. Between 50% and 79% of undergraduate pupils will commit plagiarism a minimum of once throughout their time as students, according to studies [13] [14]. Turnitin, which is a service that tracks down

plagiarism online and offers instructional feedback, opened its first office in the Philippines in March 2020. The business has been collaborating with schools and universities to comprehend the pandemic's distant evaluation demands [15].

The Smith-Waterman technique aimed at a regional sequence alignment, which looks for areas where the two sequences are most comparable. Nevertheless, the SW technique's spatial complexity and compute difficulty [16]. Sequencing readings make up the information as it is in its many forms. After read matching and quality-based cutting as part of the second analysis, a complete genomic is produced. Lastly, secondary analytics is defined as the interpretation of findings and the extraction of significant information from the data. Many algorithms and methods can be used in this final phase. These studies also serve as the basis for other applications. The tertiary analysis encompasses a variety of applications, including genomic identification and the development of a vaccine or medication [17]. The NN extracts the feature of the user for generating a rating matrix. In the first block, features are extracted and the probability score is generated for output block representation [18]. The regression problem of a content-based recommendation system makes rating predictions based on the feature of the content. The features are learned to calculate the similarity between the data items based on previously used information [19]. Clustering with one or more attributes is common for identifying different information based on similarity and correlation. The clustering methods which obtain the best grouping are k-Medoids, k-Means, Gaussian Mixtures, Hierarchical clustering, Lloyd's method, CLARA and PAM etc. [20]. The attention-gathering mechanism is a recent breakthrough in DL. The mechanism of attention has shown promising results in computer vision and a variety of NLP uses such as document sentiment classification, content summarization, named entity identification, and automated translation [21]. The key contribution of this paper is the following:

- The paper underscores the limitations of traditional identification techniques in detecting evolving plagiarism strategies, setting the stage for the need for innovative approaches.
- The study introduces a comprehensive assessment framework that considers syntactic, lexical, semantic, and structural elements, emphasizing the need for a holistic perspective.
- In response to the shortcomings of string-matching methods, the research adopts a NLP framework to enhance detection accuracy.
- The preprocessing phase is described in detail, outlining intricate steps like stemming, segmentation, tokenization, case folding, and the removal of redundant elements, which collectively refine raw text data.
- The paper highlights a pivotal aspect of the methodology: the integration of two distinct metrics within the Encoder Representation from Transformers

(E-BERT) approach, enabling a more nuanced exploration of textual similarity.

- Within the NLP realm, the combination of Deep and Shallow approaches is introduced as a lens to delve into the intricate layers of meaning within the text, revealing the potential for swift recognition of substantial revisions by Deep NLP.
- The paper introduces a novel utilization of the Waterman algorithm and an English-Spanish dictionary to enhance the process of attribute selection, improving the system's discernment of plagiarism markers.

This article is arranged in the following manner: Section II examines earlier research on prediction problems using various optimization methodologies. Section III discussed about problem statement. Section IV discusses about proposed method. Section V discusses the performance evaluation. Section VI experimental evaluation comprises mathematically developed system models. The paper is concluded in Section VII.

## II. RELATED WORKS

Patrick NyanumbaMwar et al. [22] proposed the Naive Bayes model for resume selection and classification. Based on the prediction accuracy, a homogeneous Ensemble classifier model was developed for various datasets. When compared with the original Naive Bayes Classifier, the prediction accuracy was improved.

ZhanchengRen et al. [23] developed a multi-label personality detection approach based on a neural network in which the emotional and semantic features were combined. For semantic extraction of text, sentence-level embedding was generated with Bidirectional Encoder Representation from Transformers (BERT). To estimate sentiment information, text corn analysis was invoked with a sentiment dictionary.

Ullah et al. [24] utilize machine learning, to identify software plagiarism in many programming languages. Software copying and the related issue of software plagiarism are becoming increasingly serious problems in today's society. It poses a considerable danger to the computing sector, which annually suffers significant financial losses. A customized version of the initial program may be created by the clients in different kinds of languages for programming. In addition, since every original code format may have unique grammar standards, it might be difficult to identify plagiarism in numerous forms of code sources. The study suggested a technique for multitasking language software plagiarism detection utilizing machine learning methodologies. Despite affecting the real data, characteristics are extracted from the code sources using the Principal Component Analysis. It uses factor evaluation to obtain characteristics from the information set and then transforms the principal elements into adjusted proportional fundamental elements, which are then used for forecasting assessment. Following that, the source code articles are classified by expectations using the multinomial logistic regression model implemented to these elements. It provides the logistic regression's adaptation for several class

issues. Furthermore, a paired z-test is used to assess how well the predictors performed in MLR. The information in the database is gathered in five distinct and well-known languages to conduct the investigation. Every programming language was used in two distinct examinations, Stack and binary searching.

Osman et al. [25] suggested Plagiarism is a high kind of academic rebellion that undermines the entire academic enterprise. In the past few years, several initiatives have been made to detect duplication in text documents. It is necessary to improve the methodologies that scholars have recommended for spotting copied passages, especially when conceptual analysis is required. Plagiarism is on the rise in part due to the ease with which written information may be accessed and copied on the Internet. The topic of this work is text identification of plagiarism in general. It is specifically related to technique and device detecting semantic text copying based on conceptual matching with the aid of semantic role labelling and a fuzzy inference engine. To recognize stolen semantic content, we offer essential arguments nominating strategies based on the fuzzy labelling method. The recommended technique compares text by semantically valuing each term contained in a sentence. Semantic argument construction for each sentence can benefit from semantic role labelling in several ways. To select the most important disagreements, the technique suggests nominating each argument generated by the fuzzy logic.

Hadiat et al. [26] this research aims to determine how Syntax may be used to improve the writing skills of learners in narratives and to ascertain how students perceive its usage in improving descriptive text correctness. Thirty eighth-grade kids are taking part in this particular study. The surveys, the telephone conversations, and the virtual classroom observation were used to collect the data for this study. The probability table, analyzing the content, coding, and triangulation analysis are the four methods used for analyzing data. The research shows that using Grammarly can improve the precision of producing descriptive prose. The research also reveals that the majority of students have favourable opinions of using Grammarly while writing descriptive texts because it can inspire them to improve their writing abilities, make it simple for them to identify textual errors, prevent plagiarism, and help them check their work more carefully when there are errors. To improve this work, future scholars are anticipated to perform quantitative research on related topics.

Kamble et al. [27] Plagiarism may be a situation that is expanding daily since information is developing quickly and the use of computers has grown compared to earlier times. Plagiarism is the improper use of someone else's creative work. Since it might be challenging to manually identify plagiarism, this procedure should be automated. There are several techniques available that may be used to identify plagiarism. Whereas some focus on apparent plagiarism, others focus on internal plagiarism. Processing data is a discipline that may both aid in improving the effectiveness of the procedure and assist in identifying plagiarism.

Cheers et al. [28] proposed Plagiarism within the code itself has long been a problem in postsecondary computing

teaching. Several software identification solutions have been presented to help with source code plagiarism detection. Conventional detection algorithms, nevertheless, are not resistant to ubiquitous plagiarism-hiding changes therefore can be imprecise in detecting plagiarized code from the source. This article introduces BPlag, a behavioural technique for detecting source code plagiarism. BPlag is intended to be both resistant to common plagiarism-hiding modifications and competent in detecting plagiarized code from the source. Monitoring an application's actions provides more robustness and overall accuracy since behaviour is regarded as being the least vulnerable part of a program altered by plagiarism-hiding modifications. BPlag analyzes execution behaviour via the use of symbols and describes an application in a unique graph-based style. After that, plagiarism is discovered by comparing these graphs and calculating similarity scores. BPlag is tested against five regularly used source code plagiarism detection algorithms for durability, accuracy, and efficiency.

### III. PROBLEM STATEMENT

The problem statement of this work is to improve the accuracy of plagiarism detection by implementing the Smith-Waterman algorithm and the English-Spanish dictionary technique. Plagiarism detection is a crucial task in various domains, including academia, journalism, and content creation. However, existing plagiarism detection systems may not always provide accurate results, especially when dealing with text written in different languages or when dealing with paraphrased or reworded content. By incorporating this algorithm into the plagiarism detection system, the aim is to enhance its ability to detect similarities in text, even when significant modifications have been made. Additionally, the English-Spanish dictionary technique involves utilizing a bilingual dictionary to identify similar words or phrases in both English and Spanish. This technique can be particularly useful when dealing with plagiarism across different languages, as it allows for cross-lingual comparisons and can improve the system's ability to identify instances of plagiarism. Therefore, the problem statement revolves around addressing the limitations of existing plagiarism detection systems by implementing the Smith-Waterman algorithm and the English-Spanish dictionary technique, to improve the accuracy and effectiveness of plagiarism detection, particularly when dealing with cross-lingual or rephrased content [29].

### IV. PROPOSED METHOD

This study's primary objective is to investigate the use of NLP techniques for material reprocessed detection. The theory states that a thorough analysis will find a few parallels between the original piece of writing and the modified version. A novel system containing NLP processes, comprising superficial NLP and Deep NLP, as well as more sophisticated techniques, like word2vec, is suggested to check the similarity pattern. Both the initial source material and the revised material are created entirely in English alone. The corpus-based technique is used to evaluate the system by looking at many texts from various perspectives. The use of NLP (Natural Language Processing) when used on translated texts yields more precise outcomes. Although NLP work lacks



an experimental foundation, it is suited for many sets and is motivated by past research in this area. The core components of every PD system are option selection and processing. We may generalize the text during preprocessing, and option separation reduces the overall time required for exploration to expedite the analytical phases. The aforementioned approach is used at various stages of plagiarism detection. Contrarily, certain text preparation stages employ superficial NLP techniques that are extremely straightforward and require the least amount of resources, such as lowercase, stemming lemmatization of stop word removal, and the process of tokenization. The suggested structure is broken down into six separate phases. Fig. 1 shows a flow diagram of plagiarism detection.

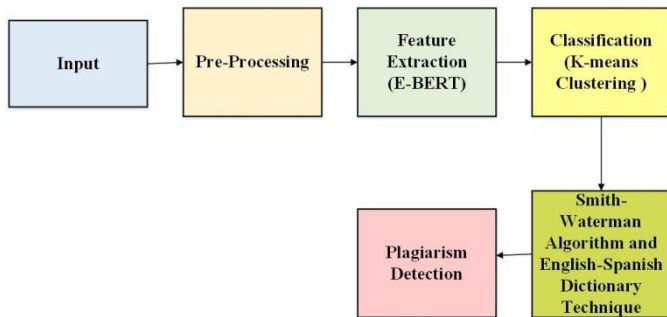


Fig. 1. Flow diagram of plagiarism detection.

#### A. Data Collection

The translations were created by qualified technical translators. For the English-Spanish language pair, the parallel corpus includes 18,303 documents, 62,057 phrases, 2,328,713 tokens that are and 14,624,745 symbols.

#### B. Pre-processing

Entering the competition and coming out on top output uses data prepared by pre-processing. Steps in preparation included eliminating stemming, segmentation, tokenization, case folding, stop word removal, null value, and special characters. This entails converting the unprocessed information into an easily readable format, which is a data mining technique by preprocessing. Data importation before using machine learning techniques is a crucial step considered by preprocessing to a textual nature being analyzed the dataset. So many steps are captured during the process. The "reviews" column and the empty rows were eliminated first. The natural language toolkit library (NLTK), a machine learning package for NLP, is also used.

The analysis yields good results, but to be sure, by spelling corrections, the meaning of the sentence has to account for sometimes spelling mistakes. The most appropriate correction is used to determine whether a word is misplaced and recommend a correction by the spellchecker. As you work with text data, the most commonly used methods are tokenization. Creating tokens from private information is the procedure to remove any unnecessary tokens, the tokenization and filtering of text data by way of sentiment analysis. With regard to sentiment analysis, stop words are words that are considered useless. In other words, removing those words won't affect the results of the model nor the precision or recall

of the analysis. They don't contribute to understanding sentences or review real significance. On very large datasets, keeping them would require higher computing power due to their size. Two methods are used to delete any stop words. Using NLTK library, the first method identified symbols with stop words and other stripped such as (e.g., a, it, is, that, and but) taken from reviews. This other method is applied to words that have a frequency greater than 50% and need to be removed from the NLTK stop words collection; use it when the word had a frequency greater than 50% but was removed as a result of low usage. Some examples are unlocked, time, mobile, and phone. Furthermore, discard the rare words that appear less than 6 times. Exclamation marks, full stops, and commas are used to remove punctuation marks. By removing both prefixes and suffixes, lemmatization or stemming returns words to their roots. By lemmas and related terms meanings are linked together. Case-folding involves replacing non-uppercase characters with their uppercase equivalents in a sequence of characters. The term "case-folding" simply refers to uppercasing when it comes to XML. Fig. 2 shows Pre-processing steps.

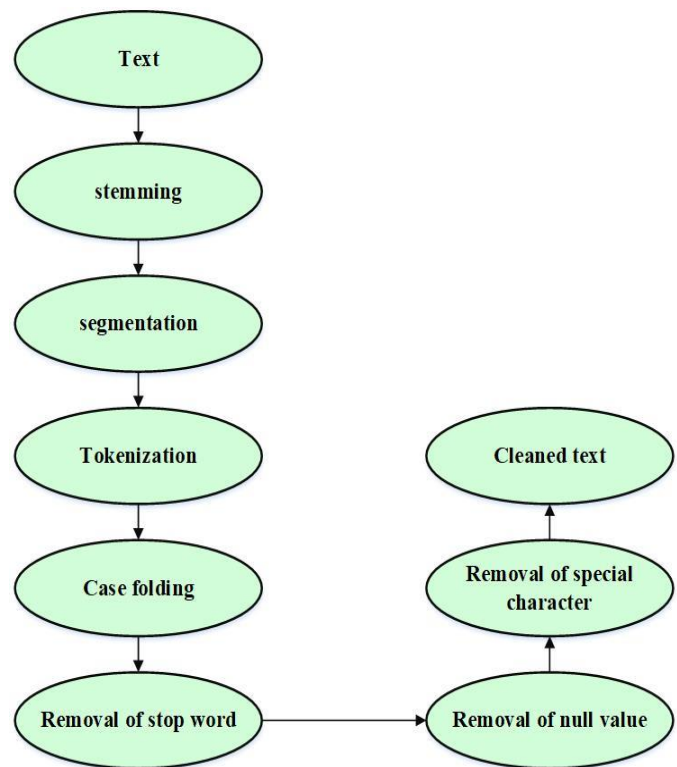


Fig. 2. Pre-processing stages.

#### C. Feature Extraction using Enhanced Bidirectional Encoder Representations from Transformers (E-BERT)

1) *Word vector*: In Chinese text, word separation does not occur and a single word is used as the text's base unit. Vectors contain information about the main features.

2) *Position vector*: Model structure alone cannot determine the placement of the input words by BERT when compared to short- and long-term memory networks and recurrent neural networks. For instance, expressing distinct

emotional dispositions using the phrases "I can't like banana chips" as "I may not like banana chips"

3) *Segment vector*: different tasks by using input and output text to meet the needs of different tasks.

Semantics-containing phrase vector in and vector output of each characters' remaining parts are shown in Fig. 3. In I-BERT, there are seven Transformer layers, of which the Encoder layer is primarily used. As part of the Encoder, attention mechanisms are used to calculate inputs and outputs and to learn features that are not possible to learn through shallow networks. Fig. 3 shows the I-BERT structure.

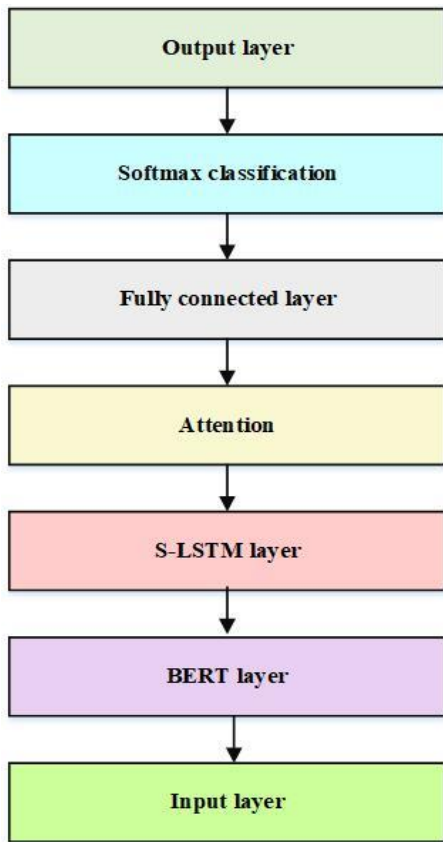


Fig. 3. I-BERT structure.

In addition to looking at the current word and obtaining semantics of the context, the self-attention mechanism does the following: incorporates a residual network and sub-layer normalization. The figure shows the structure of each transform in the I-BERT model.

Each sub-layer output is characterized as follows:

$$sub\_layer\_output = LayerNorm(x + (SubLayer(x))) \quad (1)$$

To enable information transfer between this unit's layers sublayers have been created with remaining connections. An embedded word representation makes up encoder input. An integrated feed-forward neural network is then used to process the normalized vectors. Self-attention is the main module in the Encoder section, and it is based on calculating the relationship between each word in a sentence and all of the

other words in that sentence, and then adjusting the weight of each word based on that relationship. A word vector obtained by this method includes the word's meaning but also how it interacts with other terms which makes it more global than a traditional word vector. An initialized random matrix multiplies the outputs of multiple Self-Attention mechanisms for parallel computations.

*D. Classification using K-Means Clustering (KMC) Algorithm*

Based on distance metrics, the KMC algorithm divides data samples into separate groups. It finds partitions in which the squared error between a cluster's empirical mean and its points is minimised. Let  $O = \{O_1, O_2, \dots, O_n\}$  be a set of  $n$  data samples to be clustered into a set of  $K$  clusters,  $C = C_q, q = 1, \dots, k$ . The purpose of KMC is to minimise the total of squared errors over all  $k$  clusters, which are definite as follows:

$$R(C) = \sum_q^k \sum_{O_l \in C_q} (O_l - Z_q)^2 \quad (2)$$

Where  $C_q, Z_q, O_l$  and  $k$  denote the  $q^{th}$  cluster, its centroid, data samples from the  $q^{th}$  cluster, and the total number of clusters, respectively.

Cluster centroids in KMC are generated at random. The nearest cluster to the data samples is calculated by the separations among each centroid's location and each piece of data. The average value of all the information samples within a cluster is used to modify the centre of each cluster. With the revised cluster centroids, the process of dividing the data sets into suitable clusters is then repeated until the specified termination requirements are met. Data extraction, recognition of patterns, and computer vision are just a few domains where the KMC approach has excelled. It is frequently used to give an initial setup for other sophisticated models as a pre-processing strategy [30].

Despite its benefits and popularity, KMC has some limitations because of restricted norms and effective procedures. One of the major disadvantages of KMC is its sensitivity to initialization. In particular, the method of reducing the sum of intra-cluster distances in KM is essentially a local search centred on original centroids. As a result, the initial arrangement of cluster centroids has a significant impact on KM performance optima traps. One of the primary motives for this research is the disadvantage of KMC. The process of minimizing the sum of intra-cluster distances in KMC optimized with the smith-waterman algorithm and English-Spanish dictionary technique.

$$fit(a) = \min imum(dis_{intra} + \frac{1}{dis_{inter}}) \quad (3)$$

The fitness function evaluation formula reveals that the highest efficiency is gained by lowering intra-cluster distances

and enhancing separation among clusters by maximizing inter-cluster distances [31].

### E. Smith-Waterman Algorithm and English-Spanish Dictionary Technique

In certain instances, the writing in both Spanish and English appeared to be literal translations into another language, as was seen by us. Yet, additional analytic tools have to be added to Spanish. We modified the Spanish components for tokenization when possible and sentence breaking. Use non-breaking prefixes to combine sentence breaking and tokenization, which as a result, we included in the component an inventory of Spanish non-breaking suffixes. Blocks dealing with Spanish-specific aspects were created from scratch. These cover verb tenses, comparatives, and attribute order. The position of adjectives in relation to the unit they modify is known as characteristic order. Words come after the word they modified in English; however, this is not the case in Spanish, except for a few exclusions for metaphorical effect. The element handling comparatives adds new nodes to the Spanish structure, which is particularly important in situations when there is no distinct comparable term in English. At last, a block that addresses the intricate verb tenses in Spanish was produced. This block chooses the right verb form in Spanish based on the English verb's tense, perfectiveness, and progressiveness.

Allow  $G$  as well as  $H$  stand for the patterns that need to be compatible. Let  $n$  and  $m$  stand for the lengths of  $G$  and  $H$ , accordingly. Let  $T_{q,r}$  stand for the maximum alignment score of  $G_0 \dots G_q H_0 \dots H_r$  and. Let  $U, V$  stand for the matrix to track the penalty for increasing the horizontal and vertical gaps. Let  $w(G_q, H_r)$  stand for the score of  $G_q$  aligned to  $H_r$ . The Smith-Waterman method is explained below.

$$U_{q,r} = \max \begin{cases} U_{q,r-1} - S_{ext}, \\ T_{q,r-1} - S_{first} \end{cases} \quad (4)$$

$$V_{q,r} = \max \begin{cases} V_{q-1,r} - S_{ext}, \\ T_{q-1,r} - S_{first} \end{cases} \quad (5)$$

$$T_{q,r} = \max \begin{cases} K, \\ U_{q,r}, \\ V_{q,r}, \\ T_{q-1,r-1} - w(G_q, H_r) \end{cases} \quad (6)$$

Appropriate contexts are inserted at the start and end of a statement to correspond to the words or phrases at the beginning or finish of the phrase in question. These match beacon rows and columns show a match.

## V. RESULT AND DISCUSSION

The novelty of this paper lies in its approach to plagiarism detection, particularly focusing on text and multilingual

plagiarism. The study introduces a framework that utilizes NLP methodology instead of traditional string-matching methods commonly employed for plagiarism detection. This shift in approach allows for a more comprehensive analysis of various aspects of the text, including syntactic, lexical, semantic, and structural elements. The paper also employs several pre-processing techniques, such as stemming, segmentation, tokenization, case folding, and the removal of stop words, nulls, and special characters, to prepare the text data for analysis. These steps help to improve the accuracy and effectiveness of the plagiarism detection system. This research paper introduces a novel approach to plagiarism detection by leveraging advanced NLP techniques, including E-BERT and Deep NLP. Unlike conventional methods, it integrates syntactic, lexical, semantic, and structural elements for more accurate identification. The innovative use of the Waterman algorithm and English-Spanish dictionary enhances attribute selection and captures synonym and phrase changes. This section describes the experimental setup, performance measurements, evaluation datasets, and experimental results. The proposed system will be implemented on the Python platform, and the overall performance of the proposed model will be evaluated in terms of performance metrics such as accuracy, precision, recall, specificity, and so on.

### A. Simulation Setup

An Intel(R) Core(TM) i5 processor running at 3 GHz, with four cores and four logical processors is used for the tests. The computer's name is MT, The System type is a 64-bit operating system, a 64-based processor, Microsoft Corporation is a manufacturer of operating systems, and it has built-in physical memory (RAM) of 8GB (8 GB usable).

### B. Experimental Evaluation

For performance evaluation, accuracy, precision, f-measure, recall, and Area Under the Curve (AUC) are all tested. To demonstrate the efficiency and performance of the feature learned by the suggested technique of plagiarism detection based on clustering. The proposed model is compared to models created utilizing several plagiarism encoding techniques as classifiers: word2vec+CNN, doc2vec+LR, and one-hot +LR. These techniques are supported by a variety of conditions and concepts. This study identifies the best classifier for plagiarism detection extraction.

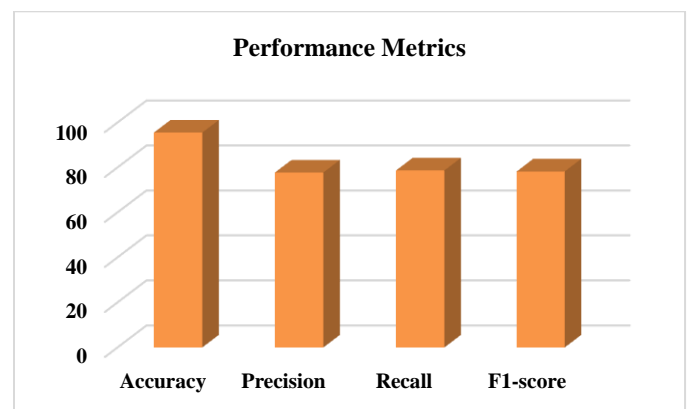


Fig. 4. Performance metrics of proposed method.

The proposed algorithm achieved a high accuracy of 99.5%, demonstrating its effectiveness. The word2vec+CNN approach achieved an accuracy of 91.18%, indicating its capability to capture semantic information. The doc2vec+LR method achieved an accuracy of 89.27%, while the one-hot encoding + logistic regression approach achieved an accuracy of 88.82%. Fig. 4 shows a comparison graph for accuracy. The proposed algorithm achieved a precision of 77.75%, indicating its ability to accurately classify positive instances. The word2vec+CNN approach achieved a precision of 59.77%, suggesting its moderate success in correctly identifying positive instances. The doc2vec+LR method achieved a precision of 51.06%, while the one-hot encoding + logistic regression approach achieved a precision of 49.19%, both demonstrating lower precision compared to the other algorithms.

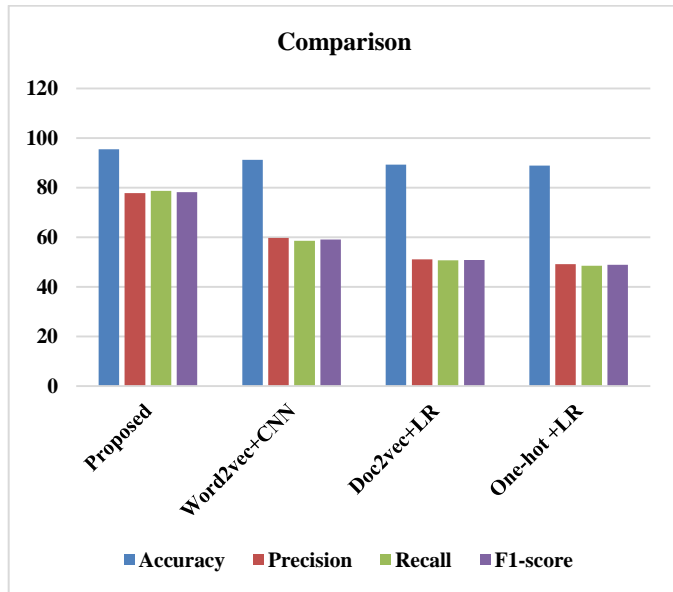


Fig. 5. Comparison graph with existing method.

Fig. 5 shows proposed algorithm achieved a high recall of 92.5%, indicating its ability to correctly identify a large proportion of positive instances. The word2vec+CNN approach achieved a recall of 58.51%, suggesting its moderate success in capturing true positive instances. The doc2vec+LR method achieved a recall of 50.67%, while the one-hot encoding + logistic regression approach achieved a recall of 48.46%, both demonstrating lower recall compared to the other algorithms. The proposed algorithm achieved a high F1-score of 98.21%, indicating its overall balance between precision and recall. The word2vec+CNN approach achieved an F1-score of 59.13%, suggesting its moderate performance in achieving a balance between precision and recall. The doc2vec+LR method achieved an F1-score of 50.87%, while the one-hot encoding + logistic regression approach achieved an F1-score of 48.82%, both demonstrating lower F1-scores compared to the other algorithms.

TABLE I. PROPOSED AND EXISTING METHODS COMPARISON

Algorithm	Accuracy	Precision	Recall	F1-score
<b>proposed</b>	95.5	77.75	78.67	78.21
<b>word2vec+CNN</b>	91.18	59.77	58.51	59.13
<b>doc2vec+LR</b>	89.27	51.06	50.67	50.87
<b>One-hot +LR</b>	88.82	49.19	48.46	48.82

Table I shows proposed algorithm achieved an accuracy of 95.5%, indicating its overall effectiveness in correctly classifying instances. It also achieved a precision of 77.75%, a recall of 78.67%, and an F1-score of 78.21%, demonstrating a good balance between precision and recall. The word2vec+CNN approach achieved a slightly lower accuracy of 91.18% with lower precision, recall, and F1-score compared to the proposed algorithm. Similarly, the doc2vec+LR and one-hot encoding + logistic regression approaches achieved lower accuracy and performance metrics compared to the proposed algorithm.

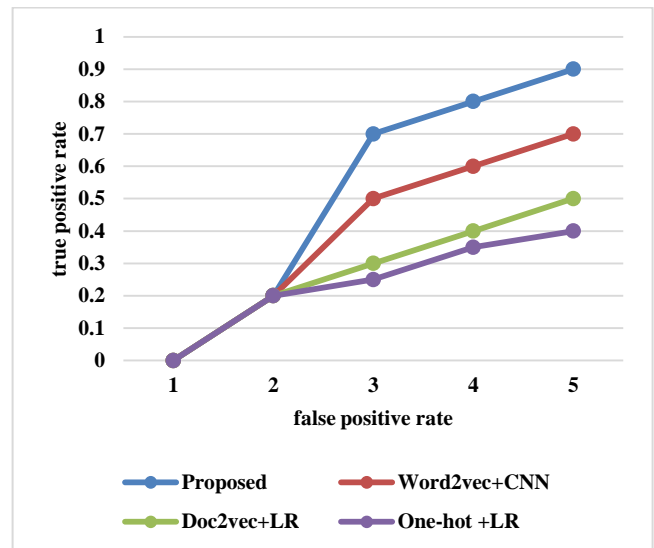


Fig. 6. AUC graph.

The above Fig. 6 shows AUC visualization was used to further analyse the performance of the suggested approach. The AUC curve has the TP rate as the y-axis and the FP rate as the x-axis with the AUC determined to indicate the models' performance. The optimal model is obtained when the AUC value is near to equal to 1.

TABLE II. AUC COMPARISON TABLE.

AUC (true positive rate)					
<b>Proposed</b>	0.1	0.2	0.7	0.8	0.9
<b>Word2vec+CNN</b>	0.1	0.2	0.5	0.6	0.7
<b>Doc2vec+LR</b>	0.1	0.2	0.3	0.4	0.5
<b>One-hot +LR</b>	0.1	0.2	0.25	0.35	0.4

The AUC Table II compares the performance of four different models across five evaluation points. The proposed model consistently achieves the highest AUC values, indicating superior predictive accuracy. The other models, including word2vec+CNN, doc2vec+LR, and One-hot+LR, demonstrate lower AUC scores, suggesting comparatively lower performance.

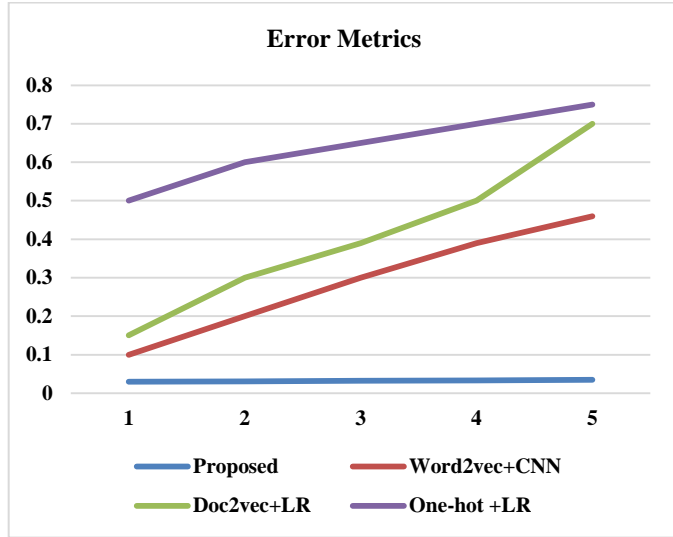


Fig. 7. Error metrics.

The error metrics consider the FPR and FNR. Fig. 7 shows the error metrics compared with existing methods. Compared with existing methods, the proposed method's error metrics are low.

$$FNR = \frac{FN}{FN + TP} = 1 - TPR \tag{7}$$

$$FPR = \frac{FP}{FP + TN} = -TNR \tag{8}$$

TABLE III. ERROR METRICS TABLE

FPR and FNR					
<b>Proposed</b>	0.03	0.031	0.032	0.033	0.035
<b>Word2vec+CNN</b>	0.1	0.2	0.3	0.39	0.46
<b>Doc2vec+LR</b>	0.15	0.3	0.39	0.5	0.7
<b>One-hot+LR</b>	0.5	0.6	0.65	0.7	0.75

Table III presents error metrics for four different models across five evaluation points. The proposed model consistently exhibits the lowest error values, indicating superior performance. Among the other models, word2vec+CNN and doc2vec+LR show intermediate error rates, while One-hot+LR has the highest error values, suggesting relatively lower accuracy.

TABLE IV. COMPARISON OF PLAGIARISM DETECTION SOFTWARE

Plagiarism Checker	Score
<b>Turnitin</b>	4.1
<b>Viper</b>	2.1
<b>Quetext</b>	2.4
<b>Proposed</b>	4.5

Table IV displays the results of several tools' plagiarism checkers. Higher numbers suggest greater possible plagiarism, with each score representing similarities or plagiarism detection levels. The greatest results go to Turnitin and Proposed, showing that they are better at spotting content similarities, while the lowest values go to Viper and Quetext, suggesting that they may be less sensitive to plagiarism. A technique for plagiarism detection can be chosen by researchers and authors depending on their own requirements.

## VI. CONCLUSION

The study focused on addressing the contemporary challenges of plagiarism detection, particularly in the context of text and multilingual plagiarism. Instead of traditional string-matching methods, an NLP methodology was employed, specifically utilizing the Encoder Representation from Transformers (E-BERT) technique. Various pre-processing techniques, such as stemming, segmentation, tokenization, case folding, and the elimination of stop words, nulls, and special characters, were applied to the text data. By integrating two measures within the E-BERT technique, the system investigated text similarity and employed the k-means clustering algorithm for categorization purposes. The deep feature representation obtained through this approach was compared to models developed using alternative encoding methods and logistic regression as a classifier, including word2vec+CNN, doc2vec+LR, and one-hot+LR. The experimental findings of the research indicated that the implemented system achieved an impressive accuracy level of 99.5% in the extraction. The utilization of the Smith-Waterman algorithm and the English-Spanish dictionary technique helped in selecting the optimal features for plagiarism detection. The future scope of this work involves advancing the plagiarism detection framework by exploring real-time, domain-specific applications and incorporating emerging transformer variants. Additionally, investigating mixed-media plagiarism detection and addressing ethical considerations for fair and transparent usage would further enhance the system's capabilities.

## REFERENCES

- [1] S. Strickroth, 'Plagiarism Detection Approaches for Simple Introductory Programming Assignments', 2021, doi: 10.18420/ABP2021-6.
- [2] D. Santos De Campos and D. James Ferreira, 'Plagiarism detection based on blinded logical test automation results and detection of textual similarity between source codes', in 2020 IEEE Frontiers in Education Conference (FIE), Uppsala, Sweden: IEEE, Oct. 2020, pp. 1-9. doi: 10.1109/FIE44824.2020.9274098.
- [3] D. Malandrino, R. De Prisco, M. Ianulardo, and R. Zaccagnino, 'An adaptive meta-heuristic for music plagiarism detection based on text similarity and clustering', Data Min. Knowl. Discov., vol. 36, no. 4, pp. 1301-1334, Jul. 2022, doi: 10.1007/s10618-022-00835-2.

- [4] M. T. J. Ansari, D. Pandey, and M. Alenezi, 'STORE: Security Threat Oriented Requirements Engineering Methodology', *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 2, pp. 191–203, Feb. 2022, doi: 10.1016/j.jksuci.2018.12.005.
- [5] M. T. J. Ansari, A. Baz, H. Alhakami, W. Alhakami, R. Kumar, and R. A. Khan, 'P-STORE: Extension of STORE Methodology to Elicit Privacy Requirements', *Arab. J. Sci. Eng.*, vol. 46, no. 9, pp. 8287–8310, Sep. 2021, doi: 10.1007/s13369-021-05476-z.
- [6] K. M. Jambi, I. H. Khan, and M. A. Siddiqui, 'Evaluation of Different Plagiarism Detection Methods: A Fuzzy MCDM Perspective', *Appl. Sci.*, vol. 12, no. 9, p. 4580, Apr. 2022, doi: 10.3390/app12094580.
- [7] M. A. Quidwai, C. Li, and P. Dube, 'Beyond Black Box AI-Generated Plagiarism Detection: From Sentence to Document Level', 2023, doi: 10.48550/ARXIV.2306.08122.
- [8] X. He, X. Shen, Z. Chen, M. Backes, and Y. Zhang, 'MGTBench: Benchmarking Machine-Generated Text Detection', 2023, doi: 10.48550/ARXIV.2303.14822.
- [9] J. Park, H. Jung, J. Lee, and J. Jo, 'An Efficient Technique of Detecting Program Plagiarism Through Program Slicing', in *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing*, R. Lee, Ed., in *Studies in Computational Intelligence*, vol. 790. Cham: Springer International Publishing, 2019, pp. 164–175. doi: 10.1007/978-3-319-98367-7\_13.
- [10] V. K and D. Gupta, 'Detection of idea plagiarism using syntax–Semantic concept extractions with genetic algorithm', *Expert Syst. Appl.*, vol. 73, pp. 11–26, May 2017, doi: 10.1016/j.eswa.2016.12.022.
- [11] I. Jarić, 'High time for a common plagiarism detection system', *Scientometrics*, vol. 106, no. 1, pp. 457–459, Jan. 2016, doi: 10.1007/s11192-015-1756-6.
- [12] M. Roostae, M. H. Sadreddini, and S. M. Fakhrahmad, 'An effective approach to candidate retrieval for cross-language plagiarism detection: A fusion of conceptual and keyword-based schemes', *Inf. Process. Manag.*, vol. 57, no. 2, p. 102150, Mar. 2020, doi: 10.1016/j.ipm.2019.102150.
- [13] H. Cheers, Y. Lin, and S. P. Smith, 'Academic Source Code Plagiarism Detection by Measuring Program Behavioral Similarity', *IEEE Access*, vol. 9, pp. 50391–50412, 2021, doi: 10.1109/ACCESS.2021.3069367.
- [14] J. Pierce and C. Zilles, 'Investigating Student Plagiarism Patterns and Correlations to Grades', in *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education*, Seattle Washington USA: ACM, Mar. 2017, pp. 471–476. doi: 10.1145/3017680.3017797.
- [15] I. K. Dhammi and R. Ul Haq, 'What is plagiarism and how to avoid it?', *Indian J. Orthop.*, vol. 50, no. 6, pp. 581–583, Dec. 2016, doi: 10.4103/0019-5413.193485.
- [16] H. Zou, S. Tang, C. Yu, H. Fu, Y. Li, and W. Tang, 'ASW: Accelerating Smith–Waterman Algorithm on Coupled CPU–GPU Architecture', *Int. J. Parallel Program.*, vol. 47, no. 3, pp. 388–402, Jun. 2019, doi: 10.1007/s10766-018-0617-3.
- [17] F. F. D. Oliveira, L. A. Dias, and M. A. C. Fernandes, 'Proposal of Smith-Waterman algorithm on FPGA to accelerate the forward and backtracking steps', *PLOS ONE*, vol. 17, no. 6, p. e0254736, Jun. 2022, doi: 10.1371/journal.pone.0254736.
- [18] R. Mishra and S. Rathi, 'Enhanced DSSM (deep semantic structure modelling) technique for job recommendation', *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 9, pp. 7790–7802, Oct. 2022, doi: 10.1016/j.jksuci.2021.07.018.
- [19] S. Benabderrahmane, N. Mellouli, and M. Lamolle, 'On the predictive analysis of behavioral massive job data using embedded clustering and deep recurrent neural networks', *Knowl.-Based Syst.*, vol. 151, pp. 95–113, Jul. 2018, doi: 10.1016/j.knosys.2018.03.025.
- [20] L. G. B. Ruiz, M. C. Pegalajar, R. Arcucci, and M. Molina-Solana, 'A time-series clustering methodology for knowledge extraction in energy consumption data', *Expert Syst. Appl.*, vol. 160, p. 113731, Dec. 2020, doi: 10.1016/j.eswa.2020.113731.
- [21] M. M. Abdelgwad, T. H. A. Soliman, A. I. Taloba, and M. F. Farghaly, 'Arabic aspect based sentiment analysis using bidirectional GRU based models', *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 9, pp. 6652–6662, Oct. 2022, doi: 10.1016/j.jksuci.2021.08.030.
- [22] P. N. Mwaro, Dr. K. Ogada, and Prof. W. Cheruiyot, 'Applicability of Naïve Bayes Model for Automatic Resume Classification', *Int. J. Comput. Appl. Technol. Res.*, vol. 9, no. 9, pp. 257–264, Sep. 2020, doi: 10.7753/IJCATR0909.1002.
- [23] Z. Ren, Q. Shen, X. Diao, and H. Xu, 'A sentiment-aware deep learning approach for personality detection from text', *Inf. Process. Manag.*, vol. 58, no. 3, p. 102532, May 2021, doi: 10.1016/j.ipm.2021.102532.
- [24] 'Software plagiarism detection in multiprogramming languages using machine learning approach - Ullah - 2021 - Concurrency and Computation: Practice and Experience - Wiley Online Library'. <https://onlinelibrary.wiley.com/doi/abs/10.1002/cpe.5000> (accessed Sep. 19, 2023).
- [25] A. H. Osman and H. M. Aljahdali, 'Important Arguments Nomination Based on Fuzzy Labeling for Recognizing Plagiarized Semantic Text', *Mathematics*, vol. 10, no. 23, p. 4613, Dec. 2022, doi: 10.3390/math10234613.
- [26] A. W. F. Hadiat, W. Tarwana, and L. Irianti, 'THE USE OF GRAMMARLY TO ENHANCE STUDENTS' ACCURACY IN WRITING DESCRIPTIVE TEXT (A CASE STUDY AT EIGHTH GRADE OF A JUNIOR HIGH SCHOOL IN CIAMIS)', vol. 9, no. 2, 2022.
- [27] S. Kamble and M. Thorat, 'CROSS-LINGUAL PLAGIARISM DETECTION USING NLP AND DATA MINING', vol. 08, no. 12, 2021.
- [28] H. Cheers, Y. Lin, and S. P. Smith, 'Academic Source Code Plagiarism Detection by Measuring Program Behavioral Similarity', *IEEE Access*, vol. 9, pp. 50391–50412, 2021, doi: 10.1109/ACCESS.2021.3069367.
- [29] T. M. Tashu, M. Lenz, and T. Horváth, 'NCC: Neural concept compression for multilingual document recommendation', *Appl. Soft Comput.*, vol. 142, p. 110348, Jul. 2023, doi: 10.1016/j.asoc.2023.110348.
- [30] H. Xie et al., 'Improving K-means clustering with enhanced Firefly Algorithms', *Appl. Soft Comput.*, vol. 84, p. 105763, Nov. 2019, doi: 10.1016/j.asoc.2019.105763.
- [31] C. Mageshkumar, S. Karthik, and V. P. Arunachalam, 'Hybrid metaheuristic algorithm for improving the efficiency of data clustering', *Clust. Comput.*, vol. 22, no. S1, pp. 435–442, Jan. 2019, doi: 10.1007/s10586-018-2242-8.

# Comparative Study of Machine Learning Algorithms for Phishing Website Detection

Kamal Omari

Department of Mathematics & Computer Science-Faculty of Sciences Ben M'sik,  
University Hassan II. Casablanca, Morocco

**Abstract**—Phishing, a prevalent online threat where attackers impersonate legitimate organizations to obtain sensitive information from victims, poses a significant cybersecurity challenge. Recent advancements in phishing detection, particularly machine learning-based methods, have shown promising results in countering these malicious attacks. In this study, we developed and compared seven machine learning models, namely Logistic Regression (LR), k-Nearest Neighbors (KNN), Support Vector Machine (SVM), Naive Bayes (NB), Decision Tree (DT), Random Forest (RF), and Gradient Boosting, to assess their efficiency in detecting phishing domains. Employing the UCI phishing domains dataset as a benchmark, we rigorously evaluated the performance of these models. Our findings indicate that the Gradient Boosting-based model, in conjunction with the Random Forest, exhibits superior performance compared to the other techniques and aligns with existing solutions in the literature. Consequently, it emerges as the most accurate and effective approach for detecting phishing domains.

**Keywords**—Phishing detection; cybersecurity; machine learning; Gradient Boosting; Random Forest

## I. INTRODUCTION

Phishing, a widespread and dangerous cyber-attack method, continues to pose significant threats in today's digital world. With the increasing reliance on online platforms, for various activities such as business, transactions and healthcare services, the risk of falling victim to phishing attacks has escalated [1]. Phishing attacks involve the deceptive acquisition of personal and sensitive information through a combination of technical deception and social engineering tactics [2, 3]. These attacks often utilize fraudulent emails or messages that appear to originate from reputable entities, tricking unsuspecting users into sharing their confidential data [2].

Despite advancements in cybersecurity that have greatly improved malware detection and reduced the presence of malware-hosting websites, combating phishing attacks remains challenging due to their social engineering nature [1, 4]. Phishing domains, in particular, exploit users' trust by directing them to counterfeit websites that closely resemble legitimate ones, leading to the compromise of sensitive information [5]. Falling victim to phishing attacks can have severe consequences, including identity theft, financial fraud, and reputational damage [1].

To address the persistent threat of phishing attacks, robust cybersecurity measures are required, and artificial intelligence (AI) has emerged as a promising approach [6]. Machine

learning (ML) algorithms, a subset of AI, offer the potential to detect and classify phishing attacks by analyzing patterns and indicators of fraudulent activity based on historical data [6]. By leveraging ML models, it becomes possible to enhance detection capabilities and accurately predict whether a webpage is a phishing site or legitimate [6].

The objective of this research paper is to compare the effectiveness of ML classification models in detecting phishing domains. By identifying the most accurate ML model among the considered algorithms, the aim is to enhance detection capabilities and mitigate the risks associated with visiting phishing websites, ultimately restoring consumer trust.

The rest of this paper is organized as follows: Section II offers insights into the algorithms used. In Section III, a review of the latest research on phishing attacks is presented. Section IV outlines the methodology employed in this study. The experimental results of our comparative study are presented and discussed in Section V. Finally, Section VI concludes the paper by summarizing the key findings and proposing avenues for future research.

## II. BACKGROUND

Various machine-learning classification methods have demonstrated their effectiveness in detecting phishing domains. Some of the prominent techniques encompass:

### A. Logistic Regression

Logistic regression is a prevalent statistical model utilized for binary classification tasks [7]. As a supervised learning algorithm, it predicts the probability of an instance belonging to a particular class. In logistic regression, the dependent variable is binary or categorical, while the independent variables can be continuous or categorical.

The primary objective of logistic regression is to determine the best-fitting logistic function that establishes the relationship between the independent variables and the probability of the binary outcome. This logistic function, also known as the sigmoid function, maps any real-valued number to a value between 0 and 1 (see Fig. 1) [8]. The resulting probability estimate is then utilized to classify the instances into their respective classes.

The logistic regression model estimates its parameters through maximum likelihood estimation, involving the optimization of the log-likelihood function [7]. Various optimization algorithms, such as gradient descent, are

commonly employed to minimize the cost function and obtain the optimal parameter values.

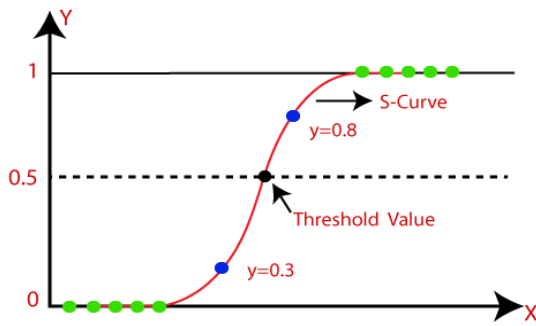


Fig. 1. Logistic regression algorithm [9].

One of the key advantages of logistic regression is its interpretability. The coefficients of the independent variables offer valuable insights into the influence and direction of each variable on the probability of the outcome [10]. Furthermore, logistic regression can effectively handle both linear and nonlinear relationships between the independent variables and the log-odds of the outcome.

### B. K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) stands as a widely employed non-parametric supervised learning algorithm for classification tasks [11]. Its simplicity, combined with its effectiveness, allows it to predict the class of an instance based on the classes of its nearest neighbors in the feature space.

In the KNN algorithm, the parameter K represents the number of nearest neighbors considered when making a prediction. To classify a new instance, KNN calculates the distances between the instance and all the training instances in the feature space (see Fig. 2). Subsequently, it identifies the K nearest neighbors based on a distance metric, such as Euclidean distance or Manhattan distance [12].

Upon identifying the K nearest neighbors, the majority class among them is assigned to the new instance. In cases of ties, a voting mechanism can resolve the class assignment. One remarkable feature of KNN is that it is a lazy learner, as it performs classification at runtime without requiring an explicit training phase [13].

KNN exhibits versatility, as it adeptly handles both binary and multi-class classification problems. Additionally, it proves to be robust in capturing complex decision boundaries and coping with noisy data [14]. Nonetheless, it is essential to carefully consider the selection of K and the distance metric, as these choices significantly impact the algorithm's performance.

### C. Support Vector Machine

Support Vector Machine (SVM) stands as a potent supervised learning algorithm extensively employed for classification and regression tasks [16]. The fundamental objective of SVM is to identify an optimal hyperplane that effectively segregates data points into distinct classes within the feature space (see Fig. 3).

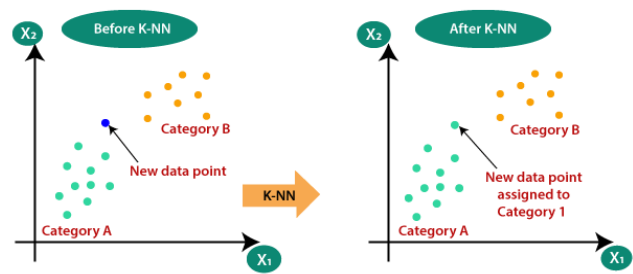


Fig. 2. K-Nearest neighbors' algorithm [15].

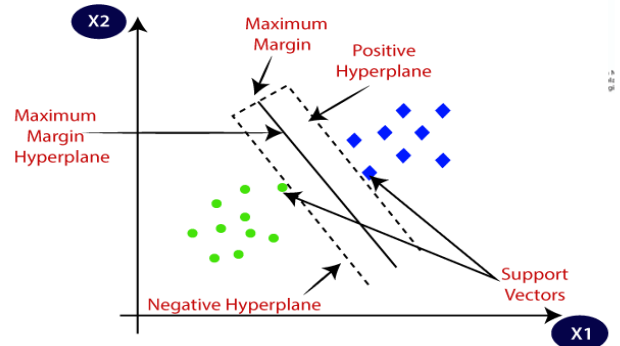


Fig. 3. Support vector machine algorithm [17].

In the SVM approach, the algorithm maps input data into a higher-dimensional feature space using a kernel function, which could be linear, polynomial, or radial basis function (RBF) kernel [18]. By transforming the data in this manner, SVM can discover a hyperplane that maximizes the margin between classes, thereby enhancing its generalization capability.

A key aspect of SVM is to identify the hyperplane that not only separates classes but also maximizes the distance to the nearest data points, known as support vectors. This property renders SVM robust to outliers and allows it to accommodate non-linear decision boundaries through various kernel functions.

SVM is versatile, accommodating both binary and multi-class classification tasks. For binary classification, SVM endeavors to locate a decision boundary that effectively distinguishes between the two classes. In multi-class scenarios, SVM can be extended using approaches such as one-vs-one or one-vs-rest [19].

### D. Naive Bayes

The Naive Bayes classifier stands as a well-known machine learning algorithm based on the application of Bayes' theorem, assuming feature independence [20]. It finds wide application in classification tasks, particularly in natural language processing and text mining.

The Naive Bayes classifier computes the probability of each class given a set of input features and selects the class with the highest probability as the predicted outcome. Its strength lies in its simplicity, efficiency, and ability to handle high-dimensional data effectively (see Fig. 4).



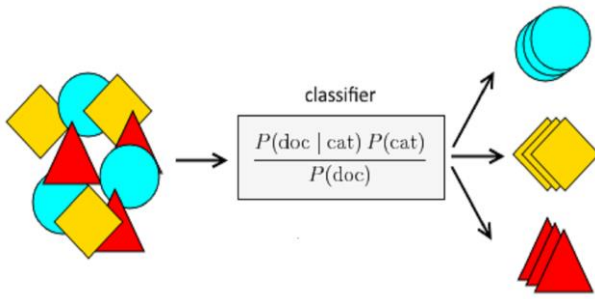


Fig. 4. Naive Bayes algorithm [21].

The algorithm is termed 'naive' due to its assumption of feature conditional independence given the class. This simplifying assumption enables efficient estimation of class probabilities by multiplying individual feature probabilities. While this assumption may not hold true in real-world scenarios, Naive Bayes often performs well and yields reliable results.

Various variations of the Naive Bayes classifier exist, such as Gaussian Naive Bayes, Multinomial Naive Bayes, and Bernoulli Naive Bayes, each suited for different data types and feature distributions [22].

Despite its simplicity, the Naive Bayes classifier has shown competitive performance compared to more complex algorithms. However, it is essential to acknowledge that Naive Bayes assumes feature independence, which may not always hold in practical scenarios.

### E. Decision Trees

Decision Trees are a well-established machine-learning algorithm utilized for classification and regression tasks [23].

These models are renowned for their intuitive and interpretable nature, constructing a tree-like flowchart based on dataset features. The Decision Tree algorithm recursively partitions the dataset, creating a tree-like structure (see Fig. 5), with internal nodes representing features and branches denoting possible feature values, leading to leaf nodes as final predicted outcomes or classes.

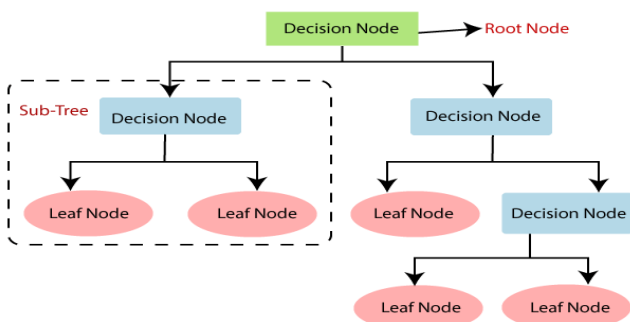


Fig. 5. Decision tree algorithm [24].

The construction involves selecting the best feature to split the data at each node based on specific criteria like information gain or Gini impurity, aiming for maximized homogeneity within subsets. Decision Trees handle both categorical and

numerical features, learning complex decision boundaries and effectively managing missing values and outliers [25].

Their applications extend to feature selection and identifying crucial features for decision-making. However, overfitting risks exist, especially with excessively deep or complex trees, mitigated by techniques like pruning and maximum tree depth setting [26].

### F. Random Forest

Random Forest stands as a widely used ensemble learning algorithm renowned for its ability to combine multiple decision trees to make accurate predictions [27]. Its versatility and robustness make it well-suited for a wide range of classification and regression tasks.

The essence of Random Forest lies in building an ensemble of decision trees, each trained on a different subset of the training data through bootstrapping [27]. Moreover, at each node of the tree, only a random subset of features is considered for splitting, introducing an element of randomness that helps mitigate overfitting and enhances the model's generalization ability (see Fig. 6).

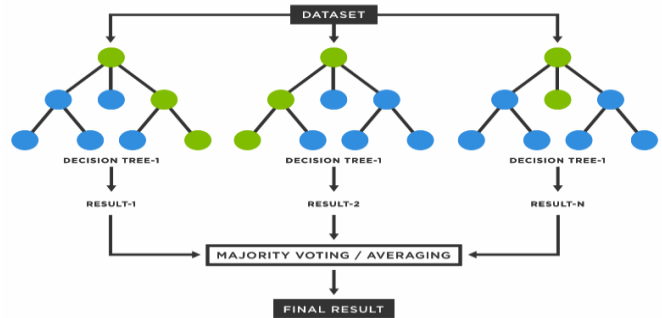


Fig. 6. Random forest algorithm [28].

To produce the final prediction, Random Forest aggregates the predictions of all individual trees either through voting (for classification) or averaging (for regression). This ensemble approach effectively reduces variance and improves the overall performance of the model.

The advantages of Random Forest are numerous, encompassing its capacity to handle high-dimensional data, automatically select important features, and remain robust in the presence of outliers and noisy data [29]. Additionally, it facilitates the estimation of feature importance, offering valuable insights into the underlying data.

However, it is essential to acknowledge that Random Forest may suffer from high computational complexity, and its results may not be as interpretable as those of single decision trees. To optimize its performance, hyperparameter tuning, including adjusting the number of trees and the maximum depth of each tree, becomes necessary [27].

### G. Gradient Boost

Gradient Boost is a highly popular and effective machine-learning algorithm widely employed in various domains due to its ability to combine weak prediction models and create a robust predictive model [30]. It belongs to the category of

boosting algorithms, which iteratively train new models to correct the errors made by previous models.

In Gradient Boost, weak models, typically decision trees, are trained in a stage-wise manner. At each stage, the model is trained on the data with a modified version of the target variable, representing the residuals or errors of the previous models. This process focuses on minimizing the errors made by the ensemble of models (see Fig. 7).

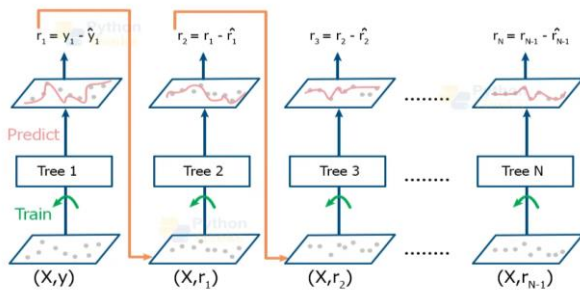


Fig. 7. Gradient boost algorithm [31].

The training process involves optimizing a loss function, such as mean squared error for regression or log loss for classification, to determine the optimal weights and parameters of the weak models. The final model is a combination of these weak models, and their predictions are weighted based on their performance on the training data.

One of the key strengths of Gradient Boost is its ability to handle complex datasets and capture non-linear relationships, leading to accurate predictions [30].

However, it is essential to be cautious about overfitting when using Gradient Boost. Controlling model complexity through regularization techniques, such as shrinkage and subsampling, can help mitigate this issue. Moreover, due to its iterative nature, Gradient Boost can be computationally expensive, necessitating careful tuning of hyperparameters, such as learning rate, tree depth, and number of iterations, to achieve optimal performance [32].

### III. RELATED WORK

URL verification is crucial for protecting users from phishing attacks, an often overlooked vulnerability [33]. However, traditional phishing detection methods exhibit limited accuracy, detecting only around 20% of attempts [33]. To overcome this, machine learning (ML) techniques have shown promise, though challenges arise with large databases and time-consuming processes [34]. Additionally, heuristics-based approaches suffer from significant false-positive rates [34]. Previous research focuses on improving anti-phishing models through feature reduction and ensemble methods [14].

Phishing URL detection is commonly treated as a classification problem using ML algorithms [35]. Constructing an ML-based detection model requires relevant properties to distinguish phishing from legitimate websites [35]. Robust ML approaches have demonstrated high detection accuracy [35], with various feature selection strategies employed to reduce feature numbers [35].

Common classifiers like Decision Trees (DT), C4.5, k-Nearest Neighbors (k-NN), and Support Vector Machines (SVM) are widely used in phishing detection research due to their accuracy and efficiency [36]. However, deep learning models face challenges such as manual parameter adjustment, lengthy training periods, and suboptimal detection accuracy [37]. Researchers emphasize the significance of ensemble learning techniques, feature selection, and reduction to address these issues [38]. Different classifiers, including Naive Bayes (NB) and SVM, have been explored [39]. Random Forest (RF) has also been successful in distinguishing phishing attacks from normal websites, with Subasi et al. [40] achieving an exceptional classification performance of 97.36% using the random forest classifier. Feature selection has been a focus in another study, with characteristics grouped to identify the most effective ones for accurate phishing attack detection [41].

In the field of phishing website detection, Patil et al. [42] proposed three strategies involving URL attribute assessment, validation based on hosting and management, and visual appearance-based analysis. They comprehensively assessed various aspects of URLs and websites using ML methodologies and algorithms.

Joshi et al. [43] conducted research on phishing attack prediction, utilizing a binary classifier based on the RF algorithm and a feature selection algorithm called relief, using data from the Mendeley domain for feature selection and training the RF algorithm.

Ubung et al. [44] explored ensemble learning strategies like bagging, boosting, and stacking to achieve high accuracy in phishing detection by integrating decision tree classifiers.

Similarly, Alsariera et al. [45] investigated phishing website detection using the "Forest by Penalizing Attributes" (FPA) algorithm and its enhanced variations, employing ensemble learning strategies like bagging, boosting, and stacking.

Pandey et al. [46] contributed to the field with a novel hybrid model combining Random Forest and Support Vector Machine (SVM) techniques for detecting phishing on websites. Their experimental results demonstrated an impressive accuracy of 94%, outperforming traditional ML algorithms SVM (90%) and Random Forest (92.96%), highlighting the superior performance of the hybrid model in classifying phishing attacks.

Furthermore, Lakshmi et al. [47] introduced an innovative approach for detecting phishing websites, analyzing hyperlinks in the HTML source code. They constructed a feature vector with 30 parameters to train a supervised DNN model with an Adam optimizer. The model demonstrated exceptional performance, outperforming traditional ML algorithms with a remarkable accuracy rate of 96%.

Table I displays a concise overview of machine learning approaches employed in phishing website detection.

TABLE I. COMPARATIVE ANALYSIS OF RECENT MACHINE LEARNING TECHNIQUES FOR PHISHING DETECTION

Model	Dataset	Algorithm	Accuracy
Subasi et al. [40]	website	RF,	97.36%
		KNN,	97.18%
		SVM,	97.17%
		ANN,	96.91%
		RF,	96.79%
		C4.5,	95.88%
		CART,	95.79%
NB	92.98%		
Patil et al.[42]	URLs	LR, DT, RF	96.23% 96.23% 96.58%
Joshi et al.[43]	Websites	RF	97.63%
Ubing et al.[44]	UCI	Ensemble bagging, boosting, stacking	95.40%
Alsariera et al. [45]	UCI	ForestPA-PWDM,	96.26%
		Bagged-ForestPA-PWDM,	96.5%
		sAdab-ForestPA-PWDM	97.4%
Pandey et al. [46]	Websites	SVM,RF	94.00%
Lakshmi et al. [47]	UCI	DNN +Adam	96.00%

IV. METHODOLOGY

In this research study, our main objective was to identify the most effective machine-learning model for detecting phishing domains. To achieve this, we conducted experiments with seven distinct machine-learning techniques: Logistic Regression (LR), k-Nearest Neighbors (KNN), Support Vector Machine (SVM), Naive Bayes (NB), Decision Tree (DT), Random Forest (RF), and Gradient Boosting.

Our dataset consisted of over 11,055 records, with 31 website parameters and a corresponding class label indicating whether it was a phishing website (1) or not (-1). To improve the models' accuracy, we applied the MinMax normalization feature as a preprocessing strategy.

To ensure robust evaluation, we employed a ten-fold cross-validation method during the classification process. This approach enabled us to obtain a more accurate performance evaluation of the models on the dataset, ensuring reliable results.

After the classification process, we thoroughly assessed the machine learning algorithms' performance using various evaluation metrics commonly used in the field, including accuracy, precision, recall, and F1-score. These metrics allowed us to make meaningful comparisons between the algorithms, ultimately identifying the most suitable approach for effectively detecting phishing websites.

To visually illustrate the concept, (see Fig. 8) presents a graphical representation of the process.

A. The Dataset

The research utilized a dataset obtained from the UCI machine-learning repository, which can be accessed at [48]. The dataset contains 11,055 records, and each sample within the dataset is composed of 31 website parameters. Among these parameters is a class label that indicates whether the website is classified as a phishing website or not, represented by values of 1 or -1 (Table II).

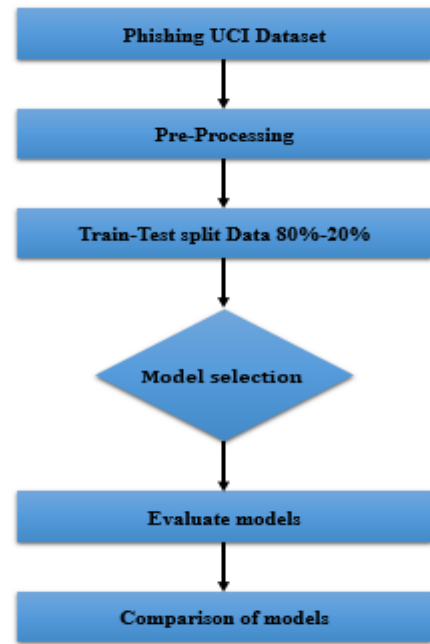


Fig. 8. Model's flowchart.

TABLE II. DESCRIPTION OF STUDIED PHISHING WEBSITE DATASET

<b>Total number of attributes</b>	31	
<b>No. of independent variables</b>	30	
<b>No. of class variables</b>	1	
<b>Details of the class variable</b>	Name: Result	
	Legitimate ==-1	Phishing=1
	4898	6175
<b>Total number of instances</b>	11055	

B. Dataset Representation

The dataset utilized in this research incorporates novel features that have been experimentally introduced [48], including the assignment of new rules to certain well-known parameters. The dataset comprises 30 parameters, which are listed below:

'having\_IP\_Address', 'URL\_Length', 'Shortening\_Service', 'having\_At\_Symbol', 'double\_slash\_redirecting', 'Prefix\_Suffix', 'having\_Sub\_Domain', 'SSLfinal\_State', 'Domain\_registration\_length', 'Favicon', 'port', 'HTTPS\_token', 'Request\_URL', 'URL\_of\_Anchor', 'Links\_in\_tags', 'SFH', 'Submitting\_to\_email', 'Abnormal\_URL', 'Redirect', 'on\_mouseover', 'RightClick', 'popUpWindow', 'Iframe', 'age\_of\_domain', 'DNSRecord', 'web\_traffic', 'Page\_Rank', 'Google\_Index', 'Links\_pointing\_to\_page', 'Statistical\_report'.

C. Visualizing the Dataset: Heatmap of Feature Correlations

To gain further insights into the dataset and understand the relationships between its features, we generated a heatmap to visualize the pairwise correlations among the 30 parameters used in this research (see Fig. 9). The heatmap provides a clear and concise representation of the correlation matrix, allowing us to identify potential associations and patterns that might influence the classification of phishing websites [50].

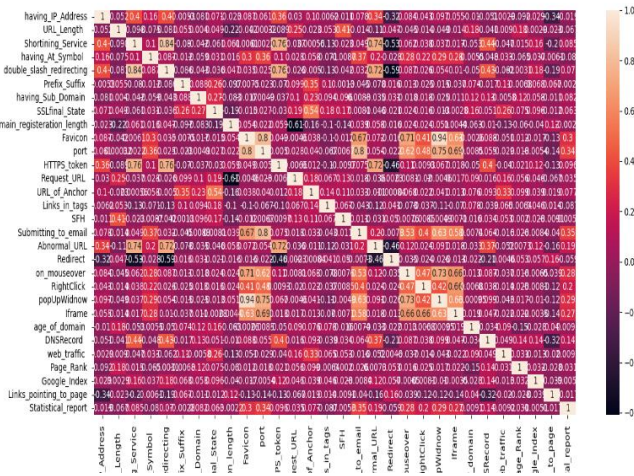


Fig. 9. The heatmap of dataset features.

Each cell in the heatmap is color-coded based on its correlation value, ranging from highly positive (dark shades) to highly negative (light shades) correlations. A correlation value close to 1 indicates a strong positive relationship, while a value close to -1 denotes a strong negative relationship. Conversely, a correlation value near 0 suggests little to no linear correlation between the respective features.

Upon analyzing the heatmap, several interesting observations emerge. For instance, we observe a strong positive correlation between the 'having\_Sub\_Domain' and 'Links\_pointing\_to\_page' features, indicating that websites with more subdomains tend to have more links pointing to their pages. Conversely, there appears to be a negative correlation between 'URL\_Length' and 'Page\_Rank', suggesting that longer URLs may be associated with lower page ranks.

Additionally, the presence of certain novel features, as experimentally introduced in the dataset, reveals potential correlations with other established parameters. For example, the 'Abnormal\_URL' feature exhibits a moderate negative correlation with 'SSLfinal\_State,' suggesting that websites with abnormal URLs might be less likely to have a valid SSL certificate.

The heatmap not only facilitates the identification of such associations but also aids in assessing potential multicollinearity among the features. Identifying multicollinearity is crucial, as it can impact the performance and interpretability of predictive models.

D. The MinMax Normalization

In our study, our main focus was to boost the precision of our proposed models by introducing MinMax normalization as a critical preprocessing measure. This technique, widely acknowledged in the realm of machine learning, significantly enhances model accuracy, particularly for specific models that rely on it [49]. By employing MinMax normalization in our suggested model, we effectively rescaled the data to a domain of [0, 1], leading to notable improvements in the input quality during model training (see Eq. (1)).

$$X_{normalized} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Where:

$X_{normalized}$  is the normalized value of the data point X.

X is the original value of the data point.

$X_{min}$  is the minimum value in the dataset.

$X_{max}$  is the maximum value in the dataset.

E. The Ten-fold Cross-validation Method

The ten-fold cross-validation method is a widely used technique in machine learning and statistical analysis to evaluate a model's performance on a dataset [51]. It involves ten iterations with different data splits for training and testing, yielding a robust estimate of the model's abilities. This approach mitigates bias and variance issues, providing a comprehensive evaluation of generalization to unseen data. The final performance metric is obtained by averaging the results from all ten iterations, ensuring an accurate assessment of the model's capabilities.

F. The Evaluation Metrics

The evaluation metrics are essential tools for assessing the performance of machine learning models. They provide quantitative measures of the model's accuracy, precision, recall, and F1-score. By using these metrics, researchers can make meaningful comparisons between different models and identify the most effective approach for their specific task.

1) Accuracy: Accuracy is a fundamental performance metric used to assess the overall correctness of a machine learning model. It represents the ratio of correctly predicted instances to the total number of instances in the dataset. In other words, it measures how often the model makes correct predictions. It is a simple and intuitive metric, but it might not be the best choice when dealing with imbalanced datasets.

$$Accuracy = \frac{\text{Number of correctly predicted instances}}{\text{Total number of instances}} \quad (2)$$

2) F1 Score: The F1 score is a balanced metric that takes into account both precision and recall. It is particularly useful when dealing with imbalanced datasets, where one class might dominate the others. The F1 score computes the harmonic mean of precision and recall, providing a single value that balances the trade-off between false positives (FP) and false negatives (FN).

$$F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (3)$$

3) Recall: Recall, also known as sensitivity or true positive rate, measures the proportion of actual positive instances that are correctly identified by the model. It is essential when the cost of false negatives is high, as it focuses on minimizing the number of missed positive instances.

$$Recall = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (4)$$

4) Precision: Precision represents the proportion of true positive predictions among all positive predictions made by the model. It is crucial when the cost of false positives is high,

as it aims to reduce the number of incorrectly classified positive instances.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \quad (5)$$

In summary, accuracy measures overall correctness, F1 score balances precision and recall, recall focuses on minimizing false negatives, and precision aims to minimize false positives.

### V. FINDINGS AND ANALYSIS

In this section, we present the experimental results of our comparative study (Table III). We analyze the performance of each machine-learning algorithm using the evaluation metrics defined in the previous section. We provide a comprehensive analysis of the results, highlighting the strengths and weaknesses of each algorithm.

TABLE III. EVALUATION RESULTS IN (%).

Classifier	Accuracy	F1 score	Recall	Precision
Gradient Boost	97.2%	96.9%	97%	96.8%
Random Forest	97.1%	97.3%	97.4%	97.2%
Decision Tree	96.3%	96.7%	96.7%	96.6%
K-Nearest Neighbors	95.6%	96.2%	96.8%	95.7%
Support Vector Machine	93.9%	95%	96.4%	93.7%
Logistic Regression	92.7%	93.8%	95%	92.7%
Naive Bayes Classifier	60.1%	45.3%	29.3%	99.2%

In this analysis, we evaluate the performance of various machine-learning models based on key metrics, including Accuracy, F1-score, Recall, and Precision. The models under consideration are Gradient Boost, Random Forest, Decision Tree, K-Nearest Neighbors, Support Vector Machine, Logistic Regression, and Naive Bayes Classifier.

Starting with Gradient Boost and Random Forest, both models showcase impressive results. Gradient Boost achieves a remarkable Accuracy of 97.2%, indicating its ability to make correct predictions for a majority of instances. The F1-score of 96.9% suggests a well-balanced trade-off between precision and recall. Additionally, with Recall and Precision scores of 97% and 96.8% respectively, it effectively identifies most positive instances while maintaining a high level of accuracy in positive predictions.

Random Forest, another strong performer, demonstrates an Accuracy of 97.1%, marginally trailing behind Gradient Boost. Nevertheless, its F1-score of 97.3% indicates an excellent balance between precision and recall. The model boasts a high Recall score of 97.4%, suggesting its proficiency in correctly identifying positive instances. Furthermore, its Precision score of 97.2% underscores its accuracy in positive predictions.

The Decision Tree model also shows promise with a respectable Accuracy of 96.3%. Its F1-score of 96.7% reflects a good balance between precision and recall. The Recall and Precision scores of 96.7% and 96.6% respectively affirm the model's effectiveness in correctly identifying positive instances and making accurate positive predictions.

K-Nearest Neighbors performs well, attaining an Accuracy of 95.6%. Its F1-score of 96.2% demonstrates a commendable balance between precision and recall. With Recall and Precision scores of 96.8% and 95.7% respectively, the model effectively identifies positive instances and makes accurate positive predictions.

The Support Vector Machine achieves an Accuracy of 93.9%, somewhat lower than the previously mentioned models. Nevertheless, its F1-score of 95% suggests a satisfactory balance between precision and recall. A Recall score of 96.4% indicates its effectiveness in identifying positive instances, and its Precision score of 93.7% underscores its accuracy in positive predictions.

Logistic Regression, with an Accuracy of 92.7%, provides a reasonable performance. Its F1-score of 93.8% signifies a good balance between precision and recall. A Recall score of 95% indicates its ability to effectively identify positive instances, while its Precision score of 92.7% reflects accurate positive predictions.

On the other hand, the Naive Bayes Classifier lags significantly behind the other models with an Accuracy of 60.1% and an F1-score of 45.3%. The low Recall score of 29.3% suggests its struggle to effectively identify positive instances. However, it exhibits an unexpectedly high Precision score of 99.2%, indicating that when it predicts a positive instance, it is usually correct. This discrepancy might imply a bias towards negative instances.

In conclusion, the analysis showcases Gradient Boost and Random Forest as top-performing models, excelling in various metrics. The Decision Tree, K-Nearest Neighbors, and Logistic Regression also demonstrate competitive performances. However, the Naive Bayes Classifier significantly underperforms in comparison to the other models, necessitating further investigation and improvements. By understanding the strengths and weaknesses of each model, we can make informed decisions when selecting the most suitable model.

After a thorough examination of our research findings, we will proceed to conduct a comparative analysis with other relevant studies in the field (Table IV).

TABLE IV. EVALUATION OF CURRENT PHISHING DOMAIN DETECTION MODELS

Authors	Dataset	Algorithm	Accuracy
Ubing et al. [44]	UCI	Ensemble bagging, boosting, stacking	95.4%
Alsariera et al. [45]	UCI	ForestPA-PWDM, Bagged-ForestPA-PWDM, and Adab-ForestPA-PWDM	96.26% 96.5% 97.4%
Lakshmi et al. [47]	UCI	DNN +Adam	96.00%
Alnemari et al. [49]	UCI	Random Forest	97.3%

Ubing et al. [44], in their work, harnessed the power of ensemble learning techniques such as bagging, boosting, and stacking, securing an impressive accuracy of 95.4% on the UCI dataset. This commendable outcome highlights the prowess of ensemble methods in accurately identifying phishing attacks, emphasizing the model's capacity to make precise predictions.

Equally noteworthy, Alsariera et al. [45] proposed multiple meta-learner models rooted in ForestPA-PWDM, Bagged-ForestPA-PWDM, and Adab-ForestPA-PWDM. Their experimental results yielded outstanding accuracies of 96.26%, 96.5%, and a remarkable 97.4%, respectively, on the UCI dataset. This exemplifies the efficacy and versatility of their ensemble-based approaches, which outperformed a majority of existing techniques, underscoring their potential as powerful solutions for phishing detection.

Venturing into the realm of deep learning, Lakshmi et al. [47] introduced the DNN +Adam model, accomplishing an impressive accuracy of 96.00% on the UCI dataset. This result vividly illustrates the effectiveness of deep learning methodologies in tackling phishing attacks, showcasing the model's ability to discern malicious websites with a remarkable degree of accuracy.

Furthermore, Alnemari et al. [49] adopted the Random Forest model, achieving an exceptional accuracy of 97.3% on the UCI dataset. The success of this approach showcases the formidable strength of Random Forest in detecting phishing attacks, operating at a high level of accuracy and outperforming several alternative methods.

Now, shifting our focus to our own research, our results reinforce the notion of competitive performance in the realm of phishing attack detection. With accuracy values ranging from 93.9% to 97.2%, the machine learning models utilized in our study prove their efficacy. Particularly, the Gradient Boost and Random Forest models demonstrated remarkable accuracies of 97.2% and 97.1%, respectively, rivaling or even surpassing the accuracy rates reported in the aforementioned studies.

The alignment of our results with the findings of previous studies accentuates the effectiveness of diverse machine learning techniques in detecting phishing attacks.

Overall, our research findings substantiate commendable performance, with the competitive accuracies achieved by our models showcasing their potential practicality in real-world phishing detection scenarios. The insights gleaned from our study empower us to make informed decisions when selecting the most appropriate machine learning technique to combat phishing threats in diverse applications.

## VI. CONCLUSION

In this research, we have delved into the critical domain of phishing website detection, exploring diverse machine learning techniques and their effectiveness in countering the ever-evolving cybersecurity threat posed by phishing attacks. The rampant increase in such fraudulent activities has presented significant challenges to individuals and organizations worldwide, necessitating the development of robust and efficient methods to detect and thwart these deceitful websites.

Through an extensive analysis of our research results and a thorough comparison with other relevant studies, we have uncovered promising insights into the effectiveness of various machine learning models for detecting phishing attacks. Notably, the Gradient Boost and Random Forest models have demonstrated exceptional performance, showcasing accuracy rates that either align with those reported in the existing

literature. This remarkable potential positions these models as viable candidates for real-world applications in phishing detection scenarios, playing a pivotal role in fortifying cybersecurity measures and shielding users from the dangers posed by phishing attacks.

## REFERENCES

- [1] Z. Alkhalil, C. Hewage, L. Nawaf, and I. Khan, "Phishing Attacks: A Recent Comprehensive Study and a New Anatomy," *Front. Comput. Sci.*, vol. 3, p. 563060, 2021, doi: 10.3389/fcomp.2021.563060.
- [2] H. Aldawood and G. Skinner, "An Advanced Taxonomy for Social Engineering Attacks," *International Journal of Computer Applications*, vol. 177, pp. 975-8887, 2020. doi: 10.5120/ijca2020919744.
- [3] R. Dhamija, J. D. Tygar, and M. Hearst, "Why phishing works," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2006, pp. 581-590.
- [4] C. Herley, "So long, and no thanks for the externalities: The rational rejection of security advice by users," in *Proceedings of the 2009 workshop on New security paradigms*, 2011, pp. 133-144.
- [5] K. L. Chiew, K. Yong, and C. C. L. Tan, "A Survey of Phishing Attacks: Their Types, Vectors and Technical Approaches," *Expert Systems with Applications*, vol. 106, 2018, pp. 10.1016/j.eswa.2018.03.050.
- [6] D. Gavrilut and I. Zaporozhan, "The use of machine learning algorithms for phishing detection," in *2019 International Conference on Innovations in Intelligent Systems and Applications (INISTA)*, 2019, pp. 1-5.
- [7] D. Hosmer and S. Lemeshow, "Applied Logistic Regression," 2004. doi: 10.1002/9781118548387.
- [8] A. Agresti, "Foundations of Linear and Generalized Linear Models," John Wiley & Sons, 2015.
- [9] "Logistic Regression in Machine Learning—Javatpoint." Available online: <https://www.javatpoint.com/logistic-regression-in-machine-learning> (accessed on 19 July 2023).
- [10] S. Menard, "Applied Logistic Regression Analysis," Sage Publications, 2002.
- [11] 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)*, Madurai, India, 2019, pp. 1255-1260, doi: 10.1109/ICCS45141.2019.9065747.
- [12] E. Rodrigues, "Combining Minkowski and Cheyshev: New Distance Proposal and Survey of Distance Metrics Using k-Nearest Neighbours Classifier," *Pattern Recognition Letters*, vol. 110, 2018, pp. 10.1016/j.patrec.2018.03.021.
- [13] P. Cunningham, M. Cord, and S. Delany, "Supervised Learning," in *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*, IGI Global, 2010, pp. 20-36. doi: 10.1007/978-3-540-75171-7\_2.
- [14] B. V. Dasarathy, "Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques," IEEE Computer Society Press, 1991.
- [15] "K-Nearest Neighbor (KNN) Algorithm for Machine Learning—Javatpoint." Available online: <https://www.javatpoint.com/logistic-regression-in-machine-learning> (accessed on 19 July 2023).
- [16] V. Kecman, "Support Vector Machines – An Introduction," in *Support Vector Machines: Theory and Applications*, 2nd ed., Springer, 2015, pp. 1-25. doi: 10.1007/10984697\_1.
- [17] "Support Vector Machine Algorithm—Javatpoint." Available online: <https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm> (accessed on 19 July 2023).
- [18] B. Schölkopf and A. J. Smola, "Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond," MIT Press, 2002.
- [19] K. Crammer and Y. Singer, "On the algorithmic implementation of multiclass kernel-based vector machines," *Journal of Machine Learning Research*, vol. 2, Dec. 2002, pp. 265-292.
- [20] P. Flach and N. Lachiche, "Naive Bayesian Classification of Structured Data," *Machine Learning*, vol. 57, 2004, pp. 233-269. doi: 10.1023/B:MACH.0000039778.69032.ab.

- [21] "Naïve Bayes Classifier from Scratch with Hands-on Examples in R." Available online: <https://insightimi.wordpress.com/2020/04/04/naive-bayes-classifier-from-scratch-with-hands-on-examples-in-r/> (accessed on 19 July 2023).
- [22] T. Almeida, J. Almeida, and A. Yamakami, "Spam filtering: How the dimensionality reduction affects the accuracy of Naive Bayes classifiers," *J. Internet Services and Applications*, vol. 1, 2011, pp. 183-200. doi: 10.1007/s13174-010-0014-7.
- [23] A. Priyam, R. Gupta, A. Rathee, and S. Srivastava, "Comparative Analysis of Decision Tree Classification Algorithms," *International Journal of Current Engineering and Technology*, vol. 3, pp. 334-337, June 2013. doi: ISSN 2277-4106.
- [24] "Machine Learning Decision Tree Classification Algorithm—Javatpoint." Available online: <https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm> (accessed on 19 July 2023).
- [25] P. Sen, M. Hajra, and M. Ghosh, "Supervised Classification Algorithms in Machine Learning: A Survey and Review," in *Proceedings of the International Conference on Advanced Computing Technologies and Applications (ICACTA)*, 2020, pp. 142-155. doi: 10.1007/978-981-13-7403-6\_11.
- [26] K. P. Murphy, "Machine Learning: A Probabilistic Perspective," MIT Press, 2012.
- [27] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5-32, 2001.
- [28] "What is a Random Forest?" Available online: <https://www.tibco.com/reference-center/what-is-a-random-forest> (accessed on 19 July 2023).
- [29] D. R. Cutler, T. C. Edwards Jr, K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler, "Random forests for classification in ecology," *Ecology*, vol. 88, no. 11, pp. 2783-2792, 2007.
- [30] Kavzoglu, T., Teke, A. Predictive Performances of Ensemble Machine Learning Algorithms in Landslide Susceptibility Mapping Using Random Forest, Extreme Gradient Boosting (XGBoost) and Natural Gradient Boosting (NGBost). *Arab J Sci Eng* 47, 7367–7385 (2022). <https://doi.org/10.1007/s13369-022-06560-8>.
- [31] "Gradient Boosting Algorithm in Machine Learning." Available online: <https://pythongeeks.org/gradient-boosting-algorithm-in-machine-learning/> (accessed on 19 July 2023).
- [32] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of Statistics*, pp. 1189–1232, 2001.
- [33] R. Dhamija, J. D. Tygar, and M. Hearst, "Why phishing works," in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, 2006, pp. 581-590.
- [34] A. A. Yavuz and H. Polat, "Phishing websites detection using machine learning techniques," *Expert Systems with Applications*, vol. 55, pp. 225-233, 2016.
- [35] M. Alazab, R. Broadhurst, and H. Chen, "Predictive data mining for combating phishing attacks," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 41, no. 3, pp. 468-479, 2011.
- [36] S. Wang, X. Jiang, W. Cui, T. Huang, B. Li, and Y. Qian, "Robust detection of web phishing using RBF based on an improved extreme learning machine," *Expert Systems with Applications*, vol. 42, no. 21, pp. 7374-7382, 2015.
- [37] M. Alazab, R. Broadhurst, and J. Slay, "PhishNet: Predictive phishing detection system," *Information Sciences*, vol. 305, pp. 65-80, 2015.
- [38] S. K. Mahanta, N. Sarma, D. Das, and A. Das, "A novel hybrid approach for phishing detection using random forest," *Procedia Computer Science*, vol. 89, pp. 120-127, 2016.
- [39] P. K. Biswas, M. I. Hossain, D. K. Bhattacharyya, M. Nasipuri, D. K. Basu, and M. Kundu, "A feature selection mechanism for phishing detection," *Applied Soft Computing*, vol. 13, no. 8, pp. 3464-3475, 2013.
- [40] A. Subasi, E. Molah, F. Almkallawi, and T. Chaudhery, "Intelligent phishing website detection using random forest classifier," in *Proceedings of the 2017 International Conference on Engineering and Computer Technology (ICECTA)*, 2017, pp. 1-5. doi: 10.1109/ICECTA.2017.8252051.
- [41] D. Salem, "On determining the most effective subset of features for detecting phishing websites," *International Journal of Computer Applications*, vol. 122, pp. 1-7, 2015. doi: 10.5120/21813-5191.
- [42] V. Patil, P. Thakkar, C. Shah, T. Bhat, and S. P. Godse, "Detection and Prevention of Phishing Websites Using Machine Learning Approach," in *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, Pune, India, 2018, pp. 1-5. doi: 10.1109/ICCUBEA.2018.8697412.
- [43] A. Joshi and Prof. T. R. Pattanshetti, "Phishing Attack Detection using Feature Selection Techniques," in *Proceedings of International Conference on Communication and Information Processing (ICCIP)* 2019.
- [44] A. Ubing, S. Kamilia, A. Abdullah, N. Zaman, and M. Supramaniam, "Phishing Website Detection: An Improved Accuracy through Feature Selection and Ensemble Learning," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, pp. 252-257, 2019.
- [45] Y. A. Alsariera, A. V. Elijah, and A. O. Balogun, "Phishing Website Detection: Forest by Penalizing Attributes Algorithm and Its Enhanced Variations," *Arab. J. Sci. Eng.*, vol. 45, pp. 10459-10470, 2020.
- [46] A. Agresti, "Foundations of Linear and Generalized Linear Models," John Wiley & Sons, 2015.
- [47] L. Lakshmi, M. P. Reddy, C. Santhaiah, and U. J. Reddy, "Smart Phishing Detection in Web Pages Using Supervised Deep Learning Classification and Optimization Technique ADAM," *Wirel. Pers. Commun.*, vol. 118, pp. 3549-3564, 2021.
- [48] UCI Machine Learning Repository, "Phishing Websites Data Set," Available online: <https://archive.ics.uci.edu/ml/datasets/phishing+websites>.
- [49] S. Alnemari and M. Alshammari, "Detecting Phishing Domains Using Machine Learning," *Applied Sciences*, vol. 13, article no. 4649, 2023. doi: 10.3390/app13084649.
- [50] K. Fu, D. Cheng, Y. Tu, and L. Zhang, "Credit Card Fraud Detection Using Convolutional Neural Networks," in *Proceedings of the 2016 International Conference on Neural Information Processing*, pp. 483-490, 2016. doi: 10.1007/978-3-319-46675-0\_53.
- [51] M. Gaag, T. Hoffman, M. Remijsen, R. Hijman, L. de Haan, B. Meijel, P. van Harten, L. Valmaggia, M. Hert, A. Cuijpers, and D. Wiersma, "The five-factor model of the Positive and Negative Syndrome Scale - II: A ten-fold cross-validation of a revised model," *Schizophrenia research*, vol. 85, pp. 280-7, 2006. doi: 10.1016/j.schres.2006.03.021.

# Study of the Impact of the Internet of Things Integration on Competition Among 3PLs

Kenza Izikki<sup>1</sup>, Mustapha Hlyal<sup>2</sup>, Aziz Ait Bassou<sup>3</sup>, Jamila El Alami<sup>4</sup>

LASTIMI Laboratory, Graduate School of Technology EST, Mohamed V University, Rabat<sup>1,3,4</sup>  
Logistics Center of Excellence, Higher School of Textile and Clothing Industries ESITH, Casablanca<sup>2</sup>

**Abstract**—The Third-Party Logistics (3PL) industry plays an important role in modern supply chains, facilitating the efficient movement of goods and optimizing logistics operations. With the advent of advanced technologies, such as the Internet of Things (IoT), automation, artificial intelligence, and data analytics, the landscape of the 3PL industry has undergone significant transformation. With their tracking ability and real time data enabling capability, IoT technologies have gained great attention from researchers and practitioners and have been widely used in the supply chain sector. This paper employs the Cournot duopoly model within the framework of game theory to investigate the profound implications of the use of IoT technology on competition and operational strategies within the 3PL sector. In this study, we construct a Cournot duopoly model focusing on the assessment of the service level of third party logistics (3PL) within the market. We consider variables such as service level and the IoT adoption rates as crucial factors influencing the behavior of these firms. Through numerical simulations we quantify the impact of the technology on the overall profitability for both firms. Our findings have demonstrated the positive impact of integrating IoT on enhancing the profits of the 3PL firms. Additionally, the IoT adoption rates and the overall IoT integration costs play a critical role in determining market equilibrium and profit distribution.

**Keywords**—Internet of things; third party logistics; game theory; cournot duopoly

## I. INTRODUCTION

In recent years, the supply chains have faced crucial changes due numerous factors such as the pandemic crisis, increasing customer demands in terms of better quality products, quicker lead times and personalized experiences. Furthermore, in light of the emergence of new international markets, the growth of e-commerce and an increasing awareness of the importance of sustainability, the supply chains operations have become greatly complex. Consequently, there is a pressing need to adopt a fresh perspective on supply chain management by embracing the shift to digitalized sustainable operations. In order to meet the ever-changing demands of customers, the adoption of a sustainable supply chain (SSC) has become imperative for companies. Over the past few decades, there has been a notable increase in the number of studies focusing on sustainable supply chain management (SSCM).

Furthermore, organizations strive to enhance their business processes by implementing technologies aligned with industry 4.0 (I4.0). This latter is revolutionizing the current industrial landscape, bringing about a significant shift in paradigms. I4.0

encompasses various technologies. Internet of Things (IoT), artificial intelligence (AI), blockchain, and the physical internet have become ubiquitous terms in describing the modern world. These digital technologies enable the generation, collection and analysis of vast volumes of data from diverse sources, including ERP systems, mobile devices, customer purchasing behavior, product lifecycle operations, global positioning systems (GPS), radio frequency identification (RFID) tracking, surveillance videos, IoT sensors etc. Incorporating these technologies into the value chain increases flexibility, efficiency, productivity and ensure a better decision-making process [1].

The Internet of Things (IoT) is a global platform that enables the connection of smart devices through the Internet. It facilitates the seamless integration of the supply chain (SC) and communication technology (ICT) infrastructure within an organization, as well as with customers and suppliers externally. Exploring the impact of the internet of things technologies on the supply chain processes have attracted the interest of many researchers [2]–[6]. IoT has proved its impact on promoting sustainability in the supply chains, through its capability to ensure visibility traceability and real time decision making.

On the other hand, the complexity of the supply chain operations, the increased costs and intense competitiveness in the global business world have led manufacturers and businesses to consider outsourcing their logistics activities to experts while focusing on their core competencies. Fast and reliable transportation has become essential, which may lead to higher transportation costs, making 3PL services necessary [7]

Third-Party Logistics (3PL) refers to the practice of engaging external companies to manage some or all of the logistics tasks for a company. 3PL firms primarily offered transportation services. However, due to the increasing competitiveness and continuous demanding customers' needs, 3PL companies have seen the obligation to incorporate a wider range of services and specializations such as now conventional and refrigerated transportation, smart warehousing [8] as well as focusing on developing their IT skills and widen their expertise [7] and providing sustainable activities and strategies [9].

Logistics requirements and expectations have evolved in recent years in view of the continuous developments in advanced technologies and the industry 4.0 revolution. Embracing cutting-edge technologies brings numerous advantages, such as increased operational efficiencies in



transportation and warehousing, as well as effective risk management. Therefore, 3PL companies ought to implement suitable information technology that aligns with their customers' needs [10].

Literature related to the 3PL market is enriched constantly, [11], [12] propose third party logistics selection frameworks using the MCDM approach, [13] explores the challenges and value of adopting the blockchain the 3PL companies. Moreover, various decision-making models for evaluating and selecting 3PL from a sustainability point of view [14]–[18]. Outsourcing logistics activities offers numerous advantages such as better flexibility and overall greater economic benefits [9], which has driven the third-party service providers market to flourish. However, the rise in 3PL companies have resulted in increased competition among them [8].

Game theory approach is one of the preferred mathematical models used in supply chain management. Considering the structure of the supply chain and their many players, game theory offers a thorough mathematical approach helping analyzing and optimizing the decision making and configurations of the supply chains [19]. The author in [20] used game theory approach to study the impact of open innovation for achieving competitive advantage. The author in [9] used game theory to investigate the pricing strategies of 3PL for a sustainable supply chain focusing on decreasing carbon emissions and delivery time. Similarly, [21] focused on pricing decisions in three different strategies considering CSR concerns and introducing a greening degree while [22] focused on investment decisions investigating who will bear the implementation cost of the IoT.

The main objective of the theoretical games in the literature focuses on pricing decision, decision to outsource and decision to investment costs. Although the use of game theory in supply chain is fairly extensive, its use in relation to 3PL, sustainability and IoT impact is under-researched. Strategy making under competitiveness remains challenging [8]. This research will consider a duopoly competition between two 3PL firms in a homogenous environment. The mathematical model will investigate the impact of the integration of the IoT technologies on the performance and quality of service of the 3PL firms.

This research investigates the following questions:

- How sensitive is the 3PL firms profit to changes in the integration rate of the IoT?
- Can IoT be added to a strategy set to enhance the competitiveness in the market of third party logistics?

The paper will be organized as followed: Section II will present a literature review of the streams of research related to our scope, Section III will describe our mathematical model, Section IV presents and discusses a numerical analysis of our model, and lastly a conclusion of our findings will be presented in Section V.

## II. LITERATURE REVIEW

Supply chains remain one of the most challenging to manage considering their complex structure. In view of

increasing customer demands; for better service, better lead times and more sustainable service, more businesses opt for outsourcing their logistics operations in order to focus on their core expertise [18]. Third party logistics providers, also known as 3PL are a well-established logistics business that offer various logistics services. They carry out numerous tasks and activities based on the need of their customer. Their main primary service consists of transportation, however warehousing, inventory management, traceability and many other supply chain activities are carried out. Many studies have pointed out the benefits of outsourcing logistics related activities to third parties. Reducing logistics related costs as well as focusing on the company's core expertise is considered the reasons companies choose to outsource to 3PL providers [17], [18]. Other benefits include better flexibility, higher logistics performance, higher quality better and strategic and operational risk management and sustainability [17]. Research has pointed out that 3PL providers are crucial to attain supply chain sustainability [7].

In these past decades, the business world has seen a drastic change in its operational and strategic levels thanks to the introduction of cutting age technologies in light of the industry 4.0 revolution. Big data analytics, IoT, Artificial intelligence, cloud technology, cybersecurity, and robotics bring important opportunities for the improvement of the supply chain performance and efficiency. The implantation of these technologies throughout the value chain leads to an increased flexibility, efficiency, productivity and better decision-making processes [1]. IoT technologies are one of the main drivers of the shift to a more connected world, enabling traceability and visibility throughout the whole value chain [3], [23].

The application of the internet of things technologies are majorly in the logistics sector [24]. Their use in different stages of the supply chain is not new, however rapid development of technologies brings new opportunities and innovative ways to enhance the logistics operation performance. RFID QR code NFC GPS and other IoT technologies have been adopted extensively achieving a transparent and efficient environment. The shift to a digitalized value chain has brought new value-added services through internet of things, automation big data AI, etc. [25].

As the third-party logistics market has expanded considerably, competition has increased. Shifting to a smart tech driven 3PL provider is a must. Researchers showed their interest in the adoption of the IoT technologies to enhance their performance and decision-making process [26]. Several works have been conducted in relation to the use of IoT in the 3PL market, [27], [28] proposed IoT enabled systems for warehousing management, while [29] proposed an IoT based just in time milk run routing system and [30] proposed an IoT based just in time milk run routing system presented a delivery system architecture for coordinating IoT infrastructure and 3PL service. IoT technologies can be used in different core processes of 3PL services enabling real time logistics, enhanced flexibility and overall improved efficiency of logistics operations. However, this field remains under-researched requiring further developments in order to take advantage of these cutting-edge technologies [31]. On the other

hand, shifting to a smart tech driven logistics remain a challenge, mainly due to cost of investments.

### III. MODEL DESCRIPTION

The SC network comprises two 3PL firms competing in a duopoly environment and multiple suppliers with 3PL as presented in Fig. 1.

We consider a market setting where  $P_1$  and  $P_2$  are two 3PL firms offering their services to  $N$  suppliers. The 3PL firms compete in the market by selecting the quantity of service that will maximize their profit. We consider the game is simultaneous and in a fixed period  $t$ .

In Cournot setting, the two firms offer the same service. The demand function represents the relationship between the total quantity of service demanded by all consumers in the market and the price at which it is offered. The inverse linear demand function is expressed as:

$$P(S) = a - b * S$$

where,  $S$  is the total service in the market,  $P(S)$  is the price of the service,  $a$  and  $b$  are positive constants relative to the market

Within the Cournot model, which examines the actions of firms competing through output quantities rather than prices, the demand function is a critical component. In this framework, each individual firm considers the output of other firms as a constant and subsequently establishes its own output level to maximize their profit.

The firms offer a number of services namely transportation, warehousing inventory management customs and compliances, technology solutions, etc.

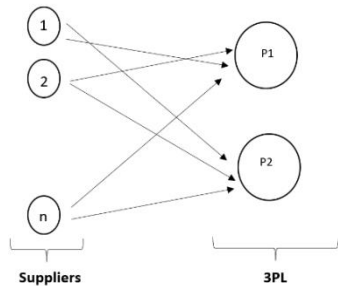


Fig. 1. The model chosen market setting comprising 3PL and suppliers.

In order to remain competitive, the firms have invested in the IoT technology to shift to a digitalized supply chain and to foster sustainability.

The use of IoT ensure a transparent and efficient value chain, greener supply chains, reduced emissions, improved lead times, and optimized costs [23], [32].

This paper considers that the firms decide to integrate the technology. We thus consider a parameter  $\beta_i \in [0,1]$  that corresponds to the integration rate of Internet of Things. The marginal cost function of a firm  $i$  is as follows:

$$C(S_i) = c * S_i + C_i^{inv} + C_i^{IoT} S_i \quad (1)$$

where

$$C_i^{inv} = \beta_i C_0 \quad (2)$$

$$C_i^{IoT} = \beta_i C_m \quad (3)$$

The marginal cost of firm  $i$  in terms of  $S_i$  and  $\beta_i$  is as follows:

$$C(S_i, \beta_i) = c * S_i + \beta_i (C_0 + C_m S_i) \quad (4)$$

Eq. (2) represents the investment cost in IoT of firm  $i$  which depends on the degree to which the technology is incorporated  $\beta_i$ , this cost is additionally adjusted by a fixed cost of investment in the IoT technology  $C_0$ . Eq. (3) corresponds to the cost of the use of IoT. It implies that the cost of the IoT use for a firm  $i$  is influenced by how extensively IoT technology is integrated in that firm  $\beta_i$ , and this cost is further scaled by the baseline cost of using IoT technology  $C_m$ .

Eq. (4) represents the marginal cost of the firms  $i$  depending on the service level  $S_i$  and their IoT integration rate  $\beta_i$ . According to the equation, when a firm haven't implemented the IoT technology its cost is equal to  $c$ , which corresponds to the basic cost of service, whereas implementing IoT induces additional charges.

The main challenge of the digitization of the supply chain and the implementation of the industry 4.0 technologies lies in their high investment cost [33]. Hence, we consider two costs related to IoT, the investment cost and the costs engendered by the exploitation of the technology such as maintenance, training, etc.

Considering the expenses associated with planning and implementing IoT initiatives in logistics, the involvement of government regulations becomes pivotal in incentivizing various industries and services [24]. We thus consider in our paper a government incentive function defined as follows:

$$I(\beta_i) = A \ln(1 + \beta_i) \quad (5)$$

where,  $A$  is constant factor that determines the scale of the incentive. In the real world, the government encouragements and incentives have a limit, thus the choice of a logarithmic function. According to (5), as the rate increases, the logarithmic component represents the diminishing returns phenomenon, where the reward gradually becomes less significant.

Based on the aforementioned, the profit function is presented as follows:

$$\Pi(S_i) = TR(S_i) - C(S_i) + I(\beta_i) S_i \quad (6)$$

Where  $TR(S_i)$  is the total revenue function for firm  $i$  defined as:  $TR(S_i) = P(S) \cdot S_i$

$$TR(S_i) = (a - b(\sum_{k \neq i} S_k + S_i)) S_i = -b S_i^2 + a S_i - b S_i \sum_{k \neq i} S_k \quad (7)$$

Substituting (4), (5) and (7) in (6), the profit equation for firm  $i$  is:

$$\Pi_i(S_i, \beta_i) = -b S_i^2 + a S_i - b S_i \sum_{k \neq i} S_k - c * S_i - \beta_i (C_0 + C_m S_i) + A \ln(1 + \beta_i) S_i \quad (8)$$

The marginal profit of the firm is expressed as follows:

$$\frac{\partial \Pi_i(S_i, \beta_i)}{\partial S_i} = -2bS_i + a - b \sum_{k \neq i} S_k - c - \beta_i C_m + A \ln(1 + \beta_i) \quad (9)$$

Since the objective function is concave as  $\frac{\partial^2 \Pi_i(S_i, \beta_i)}{\partial^2 S_i} = -2b < 0$ , the reaction function of firm  $i$  is expressed as follows considering  $\frac{\partial \Pi_i(S_i, \beta_i)}{\partial S_i} = 0$

$$S_i^*(\beta_i) = \frac{a-b \sum_{k \neq i} S_k - c}{2b} + \frac{-\beta_i C_m + A \ln(1 + \beta_i)}{2b} \quad (10)$$

This work focuses in the case of a duopoly market setting, the profit equation for firm 1 and 2 are:

$$\Pi_1(S_1, \beta_1) = -bS_1^2 + S_1[a - c - \beta_1 C_m - bS_2 + A \ln(1 + \beta_1)] - \beta_1 C_0 \quad (11)$$

$$\Pi_2(S_2, \beta_2) = -bS_2^2 + S_2[a - c - \beta_2 C_m - bS_1 + A \ln(1 + \beta_2)] - \beta_2 C_0 \quad (12)$$

And the reaction function for firm 1 and 2 are:

$$R_1(S_1) = \frac{a-bS_2-c}{2b} + \frac{-\beta_1 C_m + A \ln(1 + \beta_1)}{2b} \quad (13)$$

$$R_2(S_2) = \frac{a-bS_1-c}{2b} + \frac{-\beta_2 C_m + A \ln(1 + \beta_2)}{2b} \quad (14)$$

$$\text{Let } \begin{cases} x_1 = -\beta_1 C_m + A \ln(1 + \beta_1) \\ x_2 = -\beta_2 C_m + A \ln(1 + \beta_2) \\ a_0 = a - c \end{cases} \quad (15)$$

$S_1^*$  and  $S_2^*$  at the Cournot equilibrium are thus expressed as:

$$S_1^*(\beta_1, \beta_2) = \frac{1}{3b} (2x_1 - x_2 + a_0) \quad (16)$$

$$S_2^*(\beta_1, \beta_2) = \frac{1}{3b} (2x_2 - x_1 + a_0) \quad (17)$$

Provided that  $2x_1 - x_2 + a_0 > 0$  and  $2x_2 - x_1 + a_0 > 0$

According to this result, in the case where neither firm decides to integrate IoT;  $S_1^*(0,0) = S_2^*(0,0) = \frac{a_0}{3b}$ , both firms reach the same equilibrium output. Furthermore, if firm 1 chooses the strategy to integrate the IoT while the firm 2 chooses not to integrate, their optimal service level is:  $S_1^*(1,0) = \frac{2(-C_m + A \ln 2) + a_0}{3b}$ ,  $S_2^*(1,0) = \frac{-(-C_m + A \ln 2) + a_0}{3b}$  and  $\Delta S = \frac{-C_m + A \ln 2}{b}$ . In this case, the service level that firm 1 has to provide in this case depends on the government incentive and the cost of the use of IoT. If  $\ln 2 > C_m$ , meaning the government incentive is much greater than the IoT exploitation cost, firm 1 has to provide greater service level than firm 2. On the contrary, if the incentive is inferior to the IoT exploitation costs firm 2 should provide greater service level than firm 1.

After finding the Cournot equilibrium, it is observed that the equilibrium points depend on  $\beta_1$  and  $\beta_2$ . It shows that the integration of IoT influences the equilibrium. Therefore, in the second stage, we will analyze the profit with respect to  $\beta_1$  and  $\beta_2$ .

### A. Second Stage Equilibrium Analysis

The profit function in terms of the integration rate  $\beta_i$  by taking (16), (17) in (11), (12):

$$\Pi_1(S_1, \beta_1) = \frac{4x_1(x_1 - x_2 + a_0) + x_2(x_2 - 2a_0) + a_0^2}{9b} - \beta_1 C_0$$

$$\Pi_2(S_2, \beta_2) = \frac{4x_2(x_2 - x_1 + a_0) + x_1(x_1 - 2a_0) + a_0^2}{9b} - \beta_2 C_0$$

Thus, the profit function in terms of  $\beta_1, \beta_2$  is:

$$\begin{cases} \Pi_1(\beta_1, \beta_2) = \frac{4(-\beta_1 C_m + A \ln(1 + \beta_1))^2 + 4(-\beta_1 C_m + A \ln(1 + \beta_1))(\beta_2 C_m - A \ln(1 + \beta_2) + a_0)}{9b} + \frac{(-\beta_2 C_m + A \ln(1 + \beta_2))(-\beta_2 C_m + A \ln(1 + \beta_2) - 2a_0) + a_0^2}{9b} - \beta_1 C_0 \\ \Pi_2(\beta_1, \beta_2) = \frac{4(-\beta_2 C_m + A \ln(1 + \beta_2))^2 + 4(-\beta_2 C_m + A \ln(1 + \beta_2))(\beta_1 C_m - A \ln(1 + \beta_1) + a_0)}{9b} + \frac{(-\beta_1 C_m + A \ln(1 + \beta_1))(-\beta_1 C_m + A \ln(1 + \beta_1) - 2a_0) + a_0^2}{9b} - \beta_2 C_0 \end{cases} \quad (18)$$

The profit equation for both firms enables the calculation of the maximum local profit based on the IoT integration rate. Thus, the derivative of the system is:

$$\begin{cases} \frac{\partial \Pi_1(\beta_1, \beta_2)}{\partial \beta_1} = \frac{4}{9b} \left( \frac{A}{(1 + \beta_1)} - C_m \right) (2x_1 - x_2 + a_0) - C_0 \\ \frac{\partial \Pi_1(\beta_1, \beta_2)}{\partial \beta_2} = \frac{2}{9b} \left( \frac{A}{(1 + \beta_2)} - C_m \right) (x_2 - 2x_1 - a_0) \end{cases} \quad (19)$$

By substituting (15) in (19) our system of derivatives is expressed as:

$$\begin{cases} \frac{\partial \Pi_1(\beta_1, \beta_2)}{\partial \beta_1} = \frac{4}{9b} \left( \frac{A}{(1 + \beta_1)} - C_m \right) (-2\beta_1 C_m + 2A \ln(1 + \beta_1) + \beta_2 C_m - A \ln(1 + \beta_2) + a_0) - C_0 \\ \frac{\partial \Pi_1(\beta_1, \beta_2)}{\partial \beta_2} = \frac{2}{9b} \left( \frac{A}{(1 + \beta_2)} - C_m \right) (-\beta_2 C_m + A \ln(1 + \beta_2) + 2\beta_1 C_m - 2A \ln(1 + \beta_1) - a_0) \end{cases} \quad (20)$$

Solving  $\frac{\partial \Pi_1(\beta_1, \beta_2)}{\partial \beta_1} = \frac{\partial \Pi_1(\beta_1, \beta_2)}{\partial \beta_2} = 0$  simultaneously, the optimal equilibrium solution can be derived:

The solution for  $\frac{\partial \Pi_1(\beta_1, \beta_2)}{\partial \beta_2} = 0$  is either  $\beta_2 = \frac{A}{C_m} - 1$  or  $x_2 - 2x_1 - a_0 = 0$ . However, according to (16)  $x_2 - 2x_1 - a_0 < 0$  thus  $\beta_2 = \frac{A}{C_m} - 1$ . By replacing  $\beta_2 = \frac{A}{C_m} - 1$  in (20) we get:

$$\frac{4}{9b} \left( \frac{A}{(1 + \beta_1)} - C_m \right) \left( -2\beta_1 C_m + 2A \ln(1 + \beta_1) + \left( \frac{A}{C_m} - 1 \right) C_m - A \ln \left( \frac{A}{C_m} \right) + a_0 \right) = C_0 \quad (21)$$

Knowing  $\lim_{x \rightarrow \infty} \frac{\ln(1+x)}{1+x} = \varepsilon$ , we consider  $\ln(1 + \beta_1) = \varepsilon(1 + \beta_1)$ . Considering  $\beta_2 = \frac{A}{C_m} - 1$  and  $\beta_2 \in [0,1]$  we assume  $A \approx C_m$  since  $C_m \leq A \leq 2C_m$ . Taking the latter in consideration in (21) we get,

$$\beta_1^2 2C_m(A\varepsilon - C_m) + \beta_1 \left( \frac{8}{9b} (C_m - A\varepsilon) + C_m(1 + a_0) + C_0 \right) + C_0 - \frac{4}{9b} (2A\varepsilon + a_0) = 0 \quad (22)$$

Finding the solution for (22) we calculate the  $\Delta = \left( \frac{8}{9} (C_m - A\varepsilon) + C_m(1 + a_0) + C_0 \right)^2 + 8C_m(A\varepsilon - C_m) \left( \frac{4}{9b} (2A\varepsilon + a_0) - C_0 \right)$ . To guarantee that our equation accepts real solutions

the following condition must be met  $\Delta > 0 \Rightarrow 9bC_0 < 4(2A\varepsilon + a_0)$ . The solution is thus:

$$\beta_1 = \frac{-\left(\frac{8}{9b}(C_m - A\varepsilon) + C_m(1 + a_0) + C_0\right) \pm \sqrt{\Delta}}{4C_m(A\varepsilon - C_m)}$$

$$\beta_1 = \frac{-\left(\frac{8}{9b}(C_m - A\varepsilon) + C_m(1 + a_0) + C_0\right) \pm \sqrt{\frac{\left(\frac{8}{9b}(C_m - A\varepsilon) + C_m(1 + a_0) + C_0\right)^2 + 8C_m(A\varepsilon - C_m)\left(\frac{4}{9b}(2A\varepsilon + a_0) - C_0\right)}{4C_m(A\varepsilon - C_m)}}}{4C_m(A\varepsilon - C_m)}$$

while  $0 \leq -\left(\frac{8}{9b}(C_m - A\varepsilon) + C_m(1 + a_0) + C_0\right) \pm \sqrt{\Delta} \leq 1$

After finding the optimal equilibrium, we investigate whether the critical points are maximums by calculating the hessian matrix.

$$H(\Pi_1(\beta_1, \beta_2)) = \begin{bmatrix} R_1 & R_2 \\ R_3 & R_4 \end{bmatrix}$$

Where

$$R_1 = \frac{\partial^2 \Pi_1(\beta_1, \beta_2)}{\partial^2 \beta_1}$$

$$R_2 = \frac{\partial^2 \Pi_1(\beta_1, \beta_2)}{\partial \beta_1 \partial \beta_2}$$

$$R_3 = \frac{\partial^2 \Pi_1(\beta_1, \beta_2)}{\partial \beta_2 \partial \beta_1}$$

$$R_4 = \frac{\partial^2 \Pi_1(\beta_1, \beta_2)}{\partial^2 \beta_2}$$

$$R_1 = \frac{4}{9b} \cdot \frac{-2A^2 \ln(1 + \beta_1) + 2A^2 - A(2C_m(\beta_1 + 2) + \alpha_2 + a_0) + 2C_m^2(\beta_1 + 1)^2}{(1 + \beta_1)^2} \quad (23)$$

$$R_4 = \frac{2}{9b} \cdot \frac{A^2(\ln(1 + \beta_2) - 1) + A(C_m(\beta_2 + 2) + 2A \ln(1 + \beta_1) - 2\beta_1 C_m - a_0) + C_m^2(-\beta_2^2 - 2\beta_2 - 1)}{(1 + \beta_2)^2} \quad (24)$$

$$R_2 = \frac{4}{9b} \left(C_m - \frac{A}{\beta_2 + 1}\right) \left(\frac{A}{\beta_1 + 1} - C_m\right) \quad (25)$$

$$R_3 = \frac{4}{9b} \left(C_m - \frac{A}{\beta_1 + 1}\right) \left(\frac{A}{\beta_2 + 1} - C_m\right) \quad (26)$$

The first principal minor is  $|H_1| = \frac{\partial^2 \Pi_1(\beta_1, \beta_2)}{\partial^2 \beta_1}$

$$|H_1| = \frac{4}{9b} \cdot \frac{-2A^2 \ln(1 + \beta_1) - A(2C_m(\beta_1 + 2) - \beta_2 C_m + A \ln(1 + \beta_2) + a_0) + 2A^2 + 2C_m^2(\beta_1 + 1)^2}{(1 + \beta_1)^2} \quad (27)$$

The second principal minor or the hessian determinant  $|H_2| = R_1 R_4 - R_2 R_3$

$$|H_2| = R_1 R_4 + \frac{4}{9b} \left(C_m - \frac{A}{\beta_2 + 1}\right)^2 \left(\frac{A}{\beta_1 + 1} - C_m\right)^2 \quad (28)$$

Following the calculations, the function admits a local maximum when:

$$\begin{cases} R_1 < 0 \\ |H_2| > 0 \end{cases} \text{ therefore, these conditions should be met:}$$

$$-2A^2 \ln(1 + \beta_1) - A(2C_m(\beta_1 + 2) - \beta_2 C_m + A \ln(1 + \beta_2) + a_0) + 2A^2 + 2C_m^2(\beta_1 + 1)^2 < 0$$

$$R_1 R_4 + \frac{4}{9b} \left(C_m - \frac{A}{\beta_2 + 1}\right)^2 \left(\frac{A}{\beta_1 + 1} - C_m\right)^2 \geq 0$$

when these conditions are met, the optimal equilibrium solution for firm 1 is

$$(\beta_1, \beta_2) = \left( \frac{-\left(\frac{8}{9b}(C_m - A\varepsilon) + C_m(1 + a_0) + C_0\right) \pm \sqrt{\frac{\left(\frac{8}{9b}(C_m - A\varepsilon) + C_m(1 + a_0) + C_0\right)^2 + 8C_m(A\varepsilon - C_m)\left(\frac{4}{9b}(2A\varepsilon + a_0) - C_0\right)}{4C_m(A\varepsilon - C_m)}}}{4C_m(A\varepsilon - C_m)}, \frac{A}{C_m} - 1 \right) \quad (29)$$

Following the same approach and calculations, the optimal equilibrium solution for firm 2 is given as:

$$(\beta_1, \beta_2) = \left( \frac{A}{C_m} - 1, \frac{-\left(\frac{8}{9b}(C_m - A\varepsilon) + C_m(1 + a_0) + C_0\right) \pm \sqrt{\frac{\left(\frac{8}{9b}(C_m - A\varepsilon) + C_m(1 + a_0) + C_0\right)^2 + 8C_m(A\varepsilon - C_m)\left(\frac{4}{9b}(2A\varepsilon + a_0) - C_0\right)}{4C_m(A\varepsilon - C_m)}}}{4C_m(A\varepsilon - C_m)} \right) \quad (30)$$

#### IV. NUMERICAL ANALYSIS AND DISCUSSION

This part of the article focuses on performing a numerical analysis to investigate the impact of the parameters and the values of  $\beta_1$  and  $\beta_2$  on the profit of both firms. According to the model, we aim to assess the effects of the IoT integration rate on the profit of 3PL firms in a duopoly market, where we assume they offer identical services, at a fixed period  $t$ .

We first explored the impact of  $\beta$  on the incentives values. Fig. 2 shows the behavior of the incentive function depending on  $\beta$  and  $A$ .

According to the plot, we observe that modifications in integration rates result in proportional alterations in the incentive value. Nevertheless, the parameter  $A$  serves as a scaling factor for the incentive, influencing the magnitude of the incentive's response to changes in beta. A higher value of  $A$  magnifies the influence of beta on the incentive, whereas a smaller  $A$  diminishes this impact.

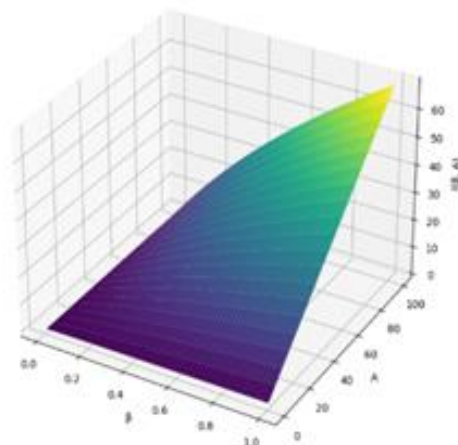


Fig. 2. Incentive function plot in terms of  $\beta$ .

We then analyzed the impact of  $\beta_1$  and  $\beta_2$  on the profit values. A data set of parameters have been examined in order to analyze the effect of the parameters and the values of  $\beta_1$  and  $\beta_2$  and on the profit values. Adjusting the values of  $C_m, A, a_0, b, C_0$  will help see how different scenarios impact the profit landscape, in different IoT integration strategies. For this purpose, we considered four different cases.

**A. Case 1: Firm 1 and Firm 2 Choose Not to Integrate IoT**

The first case we investigate is when  $\beta_1 = 0$  and  $\beta_2 = 0$ . Table I represents the value of  $\pi_1$  the profit of firm 1 and  $\pi_2$  the profit of firm 2 when both firms chose not to integrate IoT.

TABLE I. PROFIT VALUES OF FIRM 1 AND FIRM 2 AT  $\beta_1 = \beta_2 = 0$

$C_m$	$A$	$a_0$	$b$	$C_0$	$\pi_1 = \pi_2$
15	20	40	1	10	168.888889
32	60	40	1	10	168.888889
45	80	40	1	10	168.888889
15	20	80	1	10	693.333333
32	60	80	1	10	693.333333
45	80	80	1	10	693.333333
15	20	60	1	10	386.666667
32	60	60	1	10	386.666667
45	80	60	1	10	386.666667
15	20	60	1	10	386.666667
32	60	60	1	10	386.666667
45	80	60	1	20	386.666667
45	50	60	1	20	386.666667
30	50	80	1	20	693.333333
0.8	1.2	6	0.5	0.2	5.333333
1	1.5	10	0.7	0.3	12.698413

**B. Case 2 : Firm 1 and Firm 2 Integrate IoT at 100%**

The second case we consider is when both firms integrate fully the IoT. Table II represents the value of  $\pi_1$  the profit of firm 1 and  $\pi_2$  the profit of firm 2 for  $\beta_1=\beta_2=1$

The profits in this case are the same as well since they chose the same strategy of integrating IoT at 100%. In the same setting of the parameters comparing case 1 and case 2, the profits decreases due to additional costs of integrating IoT. However, in case 2 the incentives for integrating IoT impacts majorly on the profits. As the incentives increases the profits of the firms increases as well.

**C. Case 3: Firm 1 Integrate IoT at 100% and Firm 2 does not Integrate IoT**

After exploring the cases where both firms choose the same strategy, we consider the case where one firm integrates IoT fully while the other firm chooses not to integrate it. In Table III, we consider firm 1 integrates IoT at  $\beta_1=1$ , while firm 2 is at  $\beta_2=0$ .

TABLE II. PROFIT VALUES OF FIRM 1 AND FIRM 2 AT  $\beta_1 = \beta_2 = 1$

$C_m$	$A$	$a_0$	$b$	$C_0$	$\pi_1 = \pi_2$
15	20	40	1	10	138.818208
32	60	40	1	10	339.573179
45	80	40	1	10	356.835944
15	20	80	1	10	643.048317
32	60	80	1	10	1034.48573
45	80	80	1	10	1067.08971
15	20	60	1	10	346.488818
32	60	60	1	10	642.585009
45	80	60	1	10	667.58384
15	20	60	1	10	346.488818
32	60	60	1	10	642.585009
45	80	60	1	20	657.518384
45	50	60	1	20	102.748488
30	50	80	1	20	841.338431
0.8	1.2	6	0.5	0.2	5.303033
1	1.5	10	0.7	0.3	12.650859

In this case the profits are differing from firm to another. The firm with higher integration rate has greater profit since the incentives increases the profits.  $C_0$  also impacts the profits as  $C_0$  increase the profits for the firm using IoT decreases.

**D. Case 4: Both Firms Integrate IoT**

The last case we consider is both firms integrate IoT but not at a rate  $\beta=1$ . Table IV presents a number of instances with various values for integration rates for both firms 1 and 2.

TABLE III. PROFIT VALUES OF FIRM 1 AND FIRM 2 AT  $\beta_1=1$  AND  $\beta_2=0$

$C_m$	$A$	$a_0$	$b$	$C_0$	$\pi_1$	$\pi_2$
15	20	40	1	10	139.249174	169.032544
32	60	40	1	10	370.221738	179.105075
45	80	40	1	10	393.249141	181.026621
15	20	80	1	10	643.479283	693.476989
32	60	80	1	10	1065.13429	703.54952
45	80	80	1	10	1103.50291	705.471065
15	20	60	1	10	346.919784	386.810322
32	60	60	1	10	673.233567	396.882853
45	80	60	1	10	703.93158	398.804399
15	20	60	1	10	346.919784	386.810322
32	60	60	1	10	673.233567	396.882853
45	80	60	1	20	693.93158	398.804399
45	50	60	1	20	138.405228	398.552247
30	50	80	1	20	848.568762	695.743444
0.8	1.2	6	0.5	0.2	5.303706	5.333558
1	1.5	10	0.7	0.3	12.65161	12.698663

TABLE IV. PROFIT VALUES OF FIRM 1 AND FIRM 2 AT VARIOUS INTEGRATION RATES

$\beta_1$	$\beta_2$	$C_m$	$A$	$a_0$	$b$	$C_0$	$\pi_1$	$\pi_2$
0.3333	0.0123	15	20	40	1	10	179.186	169.878
0.875	0.0058	32	60	40	1	10	374.137	181.523
0.7777	0.0041	45	80	40	1	10	410.536	184.246
0.3333	0.0208	15	20	80	1	10	717.016	696.706
0.875	0.0098	32	60	80	1	10	1070.85	712.249
0.7778	0.0069	45	80	80	1	10	1130.58	714.244
0.3333	0.0179	15	20	60	1	10	403.654	388.831
0.875	0.0084	32	60	60	1	10	677.974	402.351
0.7778	0.0060	45	80	60	1	10	726.037	404.712
0.3333	0.0179	15	20	60	1	10	403.654	388.831
0.875	0.0084	32	60	60	1	10	677.974	402.351
0.77778	0.0024	45	80	60	1	20	718.872	401.968
0.1111	0.0024	45	50	60	1	20	391.622	386.941
0.6667	0.0063	30	50	80	1	20	890.362	700.780
0.5	0.8345	0.8	1.2	6	0.5	0.2	5.6977	5.48943
0.5	0.5253	1	1.5	10	0.7	0.3	13.2372	13.2282

When both integrate IoT, the rate at which each integrate influences the profits. The higher the rate, the higher is the incentive and the higher is the profits. The firm with higher rate increases their profits considerably as the incentives increases. When the firms integrate IoT at close rates, the difference between the two profits decreases.

Furthermore, in the aim to help further-examine and understand our mathematical model behavior, plotting the function provides us with valuable insights. Essentially, visually clear representation of the function facilitates assessments, insight into parameter effects and supports decision-making. Fig. 1 and Fig. 2 presents the shape of the profit function plot of firm 1 and firm 2 respectively as well as their respective maximum profit related to  $\beta_1$  and  $\beta_2$ .

Fig. 3 and Fig. 4 demonstrate that as the profit fluctuates towards the maximum, it stabilizes. As the integration rate increases, the profit increases up to the maximum point where it stabilizes.

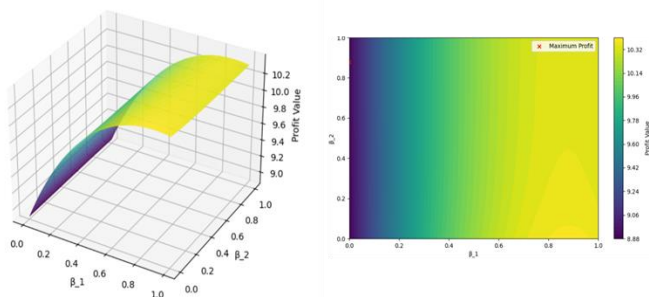


Fig. 3. Firm 1 profit visualisation.

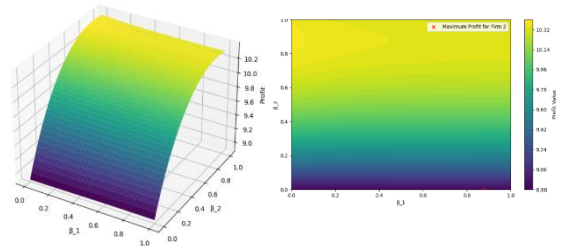


Fig. 4. Firm 2 profit visualization.

### E. Discussion

In the previous section, we explored four different cases depending on the integration rate. Based on the analysis, the four different cases we analyzed can be considered as strategies for the 3PL firms. The first strategy is when no firm integrate IoT, the second strategy is when both firms choose to fully integrate IoT, the third strategy is when one firm integrates fully the IoT while the other firms doesn't integrate it and the last strategy is when both firms integrate IoT at different rates.

Depending on the incentive that the government offers and the costs engendered by the integration of IoT, the firms can choose the strategy where they integrate IoT in order to enhance their profit. In order to be competitive, the firm needs to integrate IoT at a greater rate than its competitive firm.

Parameters such as the incentive, the cost of IoT play an important role in the profit maximization. Maximizing the profits by Integrating IoT at higher rates is contingent upon the correlation between the incentives and the costs of integrating IoT. While the incentives get higher the profits increase. If the government doesn't help considerably with the integration, it is better to choose a strategy with lower IoT rates.

### V. CONCLUSION

Globalization and the growth of e-commerce urged the logistics stakeholders to continuously strive for optimizing their operations. Logistics firms are experiencing significant pressure from customers, stakeholders and competitors to embrace digital transformation. The digitalization of the value chain has been the center of attention of both practitioner and researchers, investigating the ways to seize the opportunities from the I4.0 technologies. However, integrating these technologies is surrounded with several barriers and challenges.

Hence in this paper, we have studied the impact of integrating the IoT technologies in 3PL companies taking into consideration the high investment and maintenance costs that comes with integrating the technology. Using the Cournot duopoly model, we assessed the influence of integrating the IoT technology in the 3PL firm services on establishing competitive advantage. Our model studied the Cournot equilibrium based on the quantities of service offered by the 3PL firms in regards to the IoT integration rate. Our analysis has showed that integrating IoT can be added as a strategy set to enhance the competitiveness in the market of 3PL. Under certain conditions, integrating IoT can lead to higher profits. Both the government incentives and the IoT exploitation costs play a crucial role in determining the best strategy for the 3PL

firms. Integrating IoT at considerably higher rates than the rival firm can bring an important competitive edge, whereas integrating it at closer rates can bring greater profits for both firms.

While this study has provided valuable insights into the understanding of the 3PL industry's response to technological disruptions, it is essential to acknowledge its limitations and the broader implications of our findings. Firstly, our analysis predominantly focused on the duopoly structure, which, although prevalent in many industries, may not fully capture the complexity of the 3PL sector. Future research should explore alternative market structures, such as oligopolies or monopolistic competition, to gain a more comprehensive understanding of the dynamics at play.

Secondly, our investigation primarily focused on economic and competitive factors namely the service level and technology adoption, overlooking the growing significance of sustainability and environmental concerns in modern business environments. The influence of sustainability practices in the supply chain management industry and in the 3PL sector precisely remains a critical avenue for future exploration. Integrating sustainability considerations into our model could provide a more holistic perspective on decision-making processes within the industry.

#### REFERENCES

- [1] B. Tjahjono, C. Esplugues, E. Ares, and G. Pelaez, "What does Industry 4.0 mean to Supply Chain?," *Procedia Manuf.*, vol. 13, pp. 1175–1182, 2017. doi: 10.1016/j.promfg.2017.09.191.
- [2] N. Mostafa, W. Hamdy, and H. Alawady, "Impacts of internet of things on supply chains: A framework for warehousing," *Soc Sci.*, vol. 8, no. 3, 2019. doi: 10.3390/socsci8030084.
- [3] M. Ben-Daya, E. Hassini, and Z. Bahroun, "Internet of things and supply chain management: a literature review," *International Journal of Production Research*, vol. 57, no. 15–16. Taylor and Francis Ltd., pp. 4719–4742, 2019. doi: 10.1080/00207543.2017.1402140.
- [4] K. E. DA ROCHA, J. V. MENDES, L. A. DE SANTA-EULALIA, and V. A. D. S. MORIS, "Adoption of IoT in Logistics & Supply Chain Management: a systematic literature review," *ENEGEP 2017 - Encontro Nacional de Engenharia de Produção*, Nov. 2017. doi: 10.14488/enegep2017\_ti\_st\_238\_379\_32364.
- [5] M. Abdel-Basset, G. Manogaran, and M. Mohamed, "Internet of Things (IoT) and its impact on supply chain: A framework for building smart, secure and efficient systems," *Future Generation Computer Systems*, vol. 86. Elsevier B.V., pp. 614–628, Sep. 01, 2018. doi: 10.1016/j.future.2018.04.051.
- [6] M. Tu, "An exploratory study of internet of things (IoT) adoption intention in logistics and supply chain management a mixed research approach," in *International Journal of Logistics Management*, Emerald Group Publishing Ltd., 2018, pp. 131–151. doi: 10.1108/IJLM-11-2016-0274.
- [7] M. R. N. M. Qureshi, "A Bibliometric Analysis of Third-Party Logistics Services Providers (3PLSP) Selection for Supply Chain Strategic Advantage," *Sustainability* 2022, Vol. 14, Page 11836, vol. 14, no. 19, p. 11836, Sep. 2022. doi: 10.3390/SU141911836.
- [8] A. De and S. P. Singh, "Analysis of Competitiveness in Agri-Supply Chain Logistics Outsourcing: A B2B Contractual Framework," *Sustainability (Switzerland)*, vol. 14, no. 11, Jun. 2022. doi: 10.3390/su14116866.
- [9] M. B. Jamali and M. Rasti-Barzoki, "A game theoretic approach to investigate the effects of third-party logistics in a sustainable supply chain by reducing delivery time and carbon emissions," *J Clean Prod.*, vol. 235, pp. 636–652, Jul. 2019. doi: 10.1016/j.jclepro.2019.06.348.
- [10] J. Mageto, "Current and Future Trends of Information Technology and Sustainability in Logistics Outsourcing," *Sustainability (Switzerland)*, vol. 14, no. 13. MDPI, Jul. 01, 2022. doi: 10.3390/su14137641.
- [11] S. Boakai and F. Samanlioglu, "An MCDM approach to third party logistics provider selection," *International Journal of Logistics Systems and Management*, vol. 44, no. 3, pp. 283–299, 2023, doi: 10.1504/IJLSM.2023.129365.
- [12] B. Nila and J. Roy, "A new hybrid MCDM framework for third-party logistics provider selection under sustainability perspectives," *Expert Syst Appl.*, vol. 234, Dec. 2023. doi: 10.1016/J.ESWA.2023.121009.
- [13] Y. Zhang and N. Liu, "Blockchain adoption in serial logistics service chain: value and challenge," *Int J Prod Res.*, vol. 61, no. 13, pp. 4374–4401, 2023. doi: 10.1080/00207543.2022.2132312.
- [14] B. B. Gardas, R. D. Raut, and B. E. Narkhede, "Analysing the 3PL service provider's evaluation criteria through a sustainable approach," *International Journal of Productivity and Performance Management*, vol. 68, no. 5, pp. 958–980, Jun. 2019. doi: 10.1108/IJPPM-04-2018-0154.
- [15] H. Jung, "Evaluation of third party logistics providers considering social sustainability," *Sustainability (Switzerland)*, vol. 9, no. 5, 2017. doi: 10.3390/su9050777.
- [16] N. Kafa, Y. Hani, and A. El Mhamedi, "IFIP AICT 439 - A Fuzzy Multi Criteria Approach for Evaluating Sustainability Performance of Third - Party Reverse Logistics Providers," 2014.
- [17] K. Govindan, M. Kadziński, R. Ehling, and G. Miebs, "Selection of a sustainable third-party reverse logistics provider based on the robustness analysis of an outranking graph kernel conducted with ELECTRE I and SMAA." *Omega (United Kingdom)*, vol. 85, pp. 1–15, Jun. 2019. doi: 10.1016/j.omega.2018.05.007.
- [18] I. Dadashpour and A. Bozorgi-Amiri, "Evaluation and Ranking of Sustainable Third-party Logistics Providers using the D-Analytic Hierarchy Process," *International Journal of Engineering, Transactions B: Applications*, vol. 33, no. 11, pp. 2233–2244, Nov. 2020. doi: 10.5829/ije.2020.33.11b.15.
- [19] N. N. Vasanani, F. S. Leone Chua, L. A. Ocampo, and L. M. Brandon Pacio, "Game theory in supply chain management: current trends and applications," 2019.
- [20] A. Elmire, A. Ait Bassou, M. Hlyal, and J. El Alami, "Game Theory Approach for Open Innovation Systems Analysis in Duopolistic Market." *International Journal of Advanced Computer Science and Applications* vol. 14(5), 2023.
- [21] H. Khosroshahi, M. Rasti-Barzoki, and S. R. Hejazi, "A game theoretic approach for pricing decisions considering CSR and a new consumer satisfaction index using transparency-dependent demand in sustainable supply chains," *J Clean Prod.*, vol. 208, pp. 1065–1080, Jul. 2019. doi: 10.1016/j.jclepro.2018.10.123.
- [22] A. Dash, S. P. Sarmah, and M. K. Tiwari, "Economic Analysis of Public Priced IoT Based Traceability System in Perishable Food Supply Chain," 2022.
- [23] K. Izikki, J. El Alami, and M. Hlyal, "Internet of things in the sustainable supply chains: a systematic literature review with content analysis," *J Theor Appl Inf Technol*, vol. 15, no. 13, 2022.
- [24] H. Golpîra, S. A. R. Khan, and S. Safaeipour, "A review of logistics Internet-of-Things: Current trends and scope for future research," *J Ind Inf Integr.*, vol. 22, Jun. 2021. doi: 10.1016/j.jii.2020.100194.
- [25] B. Borgström, S. Hertz, and L. M. Jensen, "Strategic development of third-party logistics providers (TPLs): 'Going under the floor' or 'raising the roof?'," *Industrial Marketing Management*, vol. 97, pp. 183–192, Aug. 2021. doi: 10.1016/J.INDMARMAN.2021.07.008.
- [26] R. Raja and S. Venkatachalam, "Adoption of Digital Technology in Global Third-Party Logistics Services Providers: A Review of Literature," *FOCUS: Journal of International Business*, vol. 9, no. 1, pp. 105–129, 2022. doi: 10.17492/JPI.FOCUS.V9I1.912206.
- [27] W. Wu, C. Cheung, S. Y. Lo, R. Y. Zhong, and G. Q. Huang, "An IoT-enabled real-time logistics system for a third party company: A case study," *Procedia Manuf.*, vol. 49, pp. 16–23, 2020. doi: 10.1016/J.PROMFG.2020.06.005.
- [28] E. Y. C. Wong, "Development of Mobile Voice Picking and Cargo Tracing Systems with Internet of Things in Third-Party Logistics Warehouse Operations," *International Journal of Management and*

- Sustainability, vol. 5, no. 4, pp. 23–29, 2016, Accessed: May 28, 2023. [Online]. Available: <https://ideas.repec.org/a/pkp/ijomas/v5y2016i4p23-29id1020.html>
- [29] T. Qu, Y. D. Chen, Z. Z. Wang, D. X. Nie, H. Luo, and G. Q. Huang, "Internet-of-Things-based just-in-Time milk-run logistics routing system," ICNSC 2015 - 2015 IEEE 12th International Conference on Networking, Sensing and Control, pp. 258–263, Jun. 2015, doi: 10.1109/ICNSC.2015.7116045.
- [30] J. Wang, Q. Zhu, and Y. Ma, "An agent-based hybrid service delivery for coordinating internet of things and 3rd party service providers," Journal of Network and Computer Applications, vol. 36, no. 6, pp. 1684–1695, Nov. 2013, doi: 10.1016/J.JNCA.2013.04.014.
- [31] P. Evangelista, L. Santoro, and A. Thomas, "Environmental sustainability in third-party logistics service providers: A systematic literature review from 2000-2016," Sustainability (Switzerland), vol. 10, no. 5, MDPI, May 11, 2018, doi: 10.3390/su10051627.
- [32] E. Manavalan and K. Jayakrishna, "A review of Internet of Things (IoT) embedded sustainable supply chain for industry 4.0 requirements," Comput Ind Eng, vol. 127, pp. 925–953, Jul. 2019, doi: 10.1016/j.cie.2018.11.030.
- [33] A. Raj, G. Dwivedi, A. Sharma, A. B. Lopes de Sousa Jabbour, and S. Rajak, "Barriers to the adoption of industry 4.0 technologies in the manufacturing sector: An inter-country comparative perspective," Int J Prod Econ, vol. 224, p. 107546, Jun. 2020, doi: 10.1016/J.IJPE.2019.107546.



# A Framework for Predicting Academic Success using Classification Method through Filter-Based Feature Selection

Dafid<sup>1</sup>, Ermatita<sup>2</sup>, Samsuryadi<sup>3</sup>

Department of Information System, Universitas Multi Data Palembang, Palembang, Indonesia<sup>1</sup>  
Doctoral Program in Engineering Science, Universitas Sriwijaya, Palembang, Indonesia<sup>1,2,3</sup>

**Abstract**—Students’ academic success is still a serious problem faced by higher education institutions worldwide. A strategy is needed to increase the students’ academic performance and prevent students from failing. The need to get early accurate information about poor academic performance is a must and could be achieved by constructing a prediction model. Therefore, an effective technique is required to provide the accurate information and improve the accuracy of the prediction model. This study evaluates the filter-based feature selection especially the filter-based feature ranking techniques for predicting academic success. It provides a comparative study of filter-based feature selection techniques for determining the type of features (redundant, irrelevant, relevant) that affect the accuracy of the prediction models. Furthermore, this study proposes a novel feature selection technique based on attribute dependency for improving the performance of the prediction model through a framework. The experimental results show that the proposed technique significantly improved the accuracy of the prediction models from 2-8%, outperforming the existing techniques, and the Decision Tree classifier performs best for predicting with an accuracy score of 92.64%.

**Keywords**—Academic success; framework; filter-based feature selection; classifier; accuracy

## I. INTRODUCTION

Nowadays, students’ academic success is still a severe topic for researchers due to its impact on higher education quality. One of the strategies to keep and enhance it is getting early information about the poor students’ academic performance through constructing a prediction model. The ability to predict students’ academic success accurately will create opportunities to improve educational outcomes, with the result that universities can create academic policies more accurately and effectively [1].

Parallel to this, previous studies have been reported and provided varying results in creating a prediction model [1]–[3]. These reports conclude that there are two main factors in predicting academic success, which are features and prediction methods. The most common prediction method used is the classification [2]. The classification techniques frequently used by researchers to predict student academic success are Decision Tree (44%), Naive Bayes (19%), Artificial Neural Network (10%), K-Nearest Neighbor (5%), and Support Vector Machine (3%) [2]. As shown in this study [2], each technique has given the best result in educational data.

However they still suffer from accuracy because the major issue is not all of the features used in building the model hold predictive information (irrelevant and redundant features). To alleviate this limitation, researchers have applied feature selection techniques to remove the unpredicted information features.

Feature selection plays a vital role in increasing the accuracy of the prediction model. It will remove the redundant and irrelevant features and select the relevant features using a specific method [4]. Inclusion of redundant and irrelevant features will reduce the prediction model’s accuracy [5]. Each feature selection technique, namely filter, wrapper and embedded when compared to each other; filter technique is the fastest in terms of time complexity but low in performance [6]–[12]. Most researchers use filter-based feature selection on the performance of the academic success classification model due to its cheap computational cost since the academic dataset is large (more than 20 features) [13].

Further, filter-based feature selection can be divided into two types: filter-based feature ranking techniques and filter-based feature subset techniques [5], [13]–[16]. Filter-based feature ranking techniques select the features by choosing only top-ranked features using a ranker search method with important scores such as ChiSquared (Statistics-Based Scores), InfoGain and GainRatio (Probability-Based Scores), Symmetrical Uncertainty (Probability-Based Scores) and Relief (Instance-Based Techniques). In contrast, filter-based feature subset techniques can evaluate the feature subset as a whole instead of evaluating individual features using a feature-subset evaluator. Several previous research have compared and proved that filter-based feature selection gives an impact on increasing the performance of academic success classification [17]–[22]. However, the problem arises when using filter-based feature ranking techniques are computationally cheap but do not deal with redundant features [16]. These techniques tend to select redundant and irrelevant features as they do not consider feature interactions [5], [13]–[15], [23]. In contrast, filter-based feature subset techniques are better for deselecting redundant features but are computationally expensive due to repeating steps in a greedy forward fashion to find the best subset based on certain criteria [11], [16] until the desired subset is reached. Even though in these techniques there are efforts to remove redundant features by finding the relevancy between the individual features, but in the fact the techniques only discard a few redundant variables since the measures evaluate features

individually [13], [23]. A brief overview of the filter-based feature selection can be seen in Table I.

Due to these above-said problems, the main contribution of this study is a framework that can act as a guide to build the prediction model by using the classification method through filter-based feature selection that fits into the problem to enhance accuracy. This research focuses on improving filter-based feature ranking techniques by addressing redundant and irrelevant problems with attribute dependency techniques and unique attribute technique. These techniques have ability to find the relationship among features not only individually but also in groups that are related to redundancy and irrelevancy problems [24]–[26]. The evaluation measurement of accuracy is then used to test each prediction model that has been constructed.

Through the proposed framework, higher education can accurately identify failed-risk students who need special attention to prevent them from failing and take appropriate action at right time. This finding can be employed to help poor academic performance students in their next semester examination to get better results. Consequently, it will have an academic impact on higher education to enhance students' academic success.

TABLE I. OVERVIEW OF THE STUDIED FILTER-BASED SELECTION TECHNIQUES ON EDUCATIONAL DATA

Family	Method	References	Merit	Demerit
Filter-based feature ranking techniques	ChiSquared (CHI)	[20]–[22]	Redundant features are selected	Low computational cost
	InfoGain (IG)	[17], [18], [20]–[22]		
	GainRatio (GR)	[17], [18], [20]–[22]		
	Symmetrical Uncertainty (SU)	[17], [18], [22]		
	Relief (RF)	[17], [18], [20]		
Filter-based feature subset techniques	Correlation (CFS)	[17], [18], [20], [21]	Reduce redundant features	High computational cost

The following is how this section of the paper is structured: Section II reviews recent similar works on predicting academic success and limitations of them and presents feature selection methods used in this study. In Section III, the research methodology was described. The research results are given in Section IV, along with the study implications. Finally, Section V presents the conclusion in its entirety.

## II. RELATED WORK

### A. Previous Research

This section discusses the previous works related to the academic success field that have used filter-based feature selection techniques to enhance the performance of the prediction models on education data.

A research by Khasanah [17] discovered that filter-based feature selection may be used to carefully choose high influence features to predict student performance. The filter-based feature selections used are correlation-based attribute

evaluation, gain-ratio attribute evaluation, information-gain attribute evaluation, relief attribute evaluation, and symmetrical uncertainty attribute evaluation. Student attendance and first-semester GPA were shown to be the most important factors. They employed Bayesian Network and Decision Tree algorithms to classify and predict student performance, where the Bayesian Network outperformed the Decision Tree classification in the accuracy case. Hussain [18] conducted a study to find high influence attributes using correlation-based attribute evaluation, gain-ratio attribute evaluation, information-gain attribute evaluation, relief attribute evaluation, and symmetrical uncertainty attribute evaluation. The techniques showed that internal assessment attributes as highly influential attributes, where random forest outperformed the other classifiers in the accuracy case. A study was carried out by Priyasadie (19) to forecast junior high school students' grades using feature selection to enhance classification performance. The outcome demonstrated that Decision Tree outperforms other classification algorithms in general when parameter optimization and feature selection are used and First Semester Natural Science and First Semester Social Science are the most important attribute. Zaffar [20] conducted a study of feature selection techniques to improve the accuracy of academic achievement prediction. The correlation-based feature evaluator (CFS), the chi-squared test, the filtered, gain ratio (GR), the principal component analysis (PC), and the Relief method are the feature selection approaches used in the study. The study used fifteen prediction models and compared them to each other in order to verify the effectiveness of the proposed feature selection techniques. The experiment shows that using feature selection increases accuracy. Sokkhey [22] developed the CHIMI feature selection technique, a combination of the ChiSquare and Mutual Information ranking algorithms to identify the most relevant features affecting classification performance. The results presented that the technique improved the prediction model accuracy.

According to the review previous research above, each filter-based techniques have its advantage, but in general the disadvantages of them are inability or not optimal to remove the redundant and irrelevant features since they choose only top-ranked features without considering feature interactions [5], [8], [15], [23], [27]–[29].

### B. Current Research

This section presents the current work to overcome the limitations in the previous research by proposing a framework through a novel filter-based feature ranking techniques utilizing attribute dependency theory. It presents the central concepts and definitions related to feature selection.

This study presents an analysis of filter-based feature selection techniques on a number of classifiers and then determines each classifier's performance in terms of accuracy. It focuses on the existing filter-based feature ranking techniques compared to this proposed technique to justify its better performance in terms of removal of redundant and irrelevant features and improved classification accuracy, while filter-based feature subset techniques are still used to complete the result of this research.

As known, the existing technique failed to remove the redundant and irrelevant attributes because they are guided by a statistical evaluation metric that ranks the attributes in decreasing order of importance without considering the relationship between the features. Thus, highly relevant features with the class, but redundant with others, will be on the top of the ranking. Meanwhile, the attribute dependency theory has the ability to determine relationship among two or more (group) features based on dependency measures [24]–[26]. Thus, there is a chance and suitable for this theory to address the limitation of removing redundant and irrelevant attributes in the previous technique. For an effective process, the technique requires domain expert/domain knowledge that will be helpful to identify the suspect redundant features for next determining them as redundant features or not using dependency measure.

In general, features in a dataset are classified into redundant, irrelevant and relevant features (strong and or weak features) [30]. Redundant features are those that contain predictive information but it has already been conveyed through any of another features or subset of features. Irrelevant features are those that do not convey any predictive information whereas the relevant features are those that hold predictive information in building learning models. Redundancy among the features or group of features is identified based on the relationship among the features, whereas the relevance is identified by finding the relationship between the feature (or group of features) and the class.

Thus, this study considers attribute dependency to remove redundant features. Using this technique, redundant features are identified and removed by ensuring the strong dependency among the features or group of features. In contrast, the relevant features are identified and selected between the features or group of features and the class. Irrelevant attributes are identified and removed by ensuring the uniqueness values or single value (or dominant only one value).

The proposed feature selection technique is rooted on the concepts of relevance, redundancy and features interaction analysis. In this regard, preliminary concepts related to attribute dependency theory are presented here. Further, the rationale for incorporating these concepts in the proposed framework is explained. Aside from how attributes are evaluated, another important aspect is how the feature selection technique relates to the classification method that will be used to further build the model containing patterns from the selected features. In this sense, five main classifiers may be mentioned. The proposed framework of this research is shown in Fig. 1.

### C. Unique Attribute to Process the Feature Selection as Proposed Method

An attribute that has a different value for each row or record in dataset [24]–[26]. In addition, the attributes (features) that only have one value or one value dominates other values, they also are considered as unique attribute.

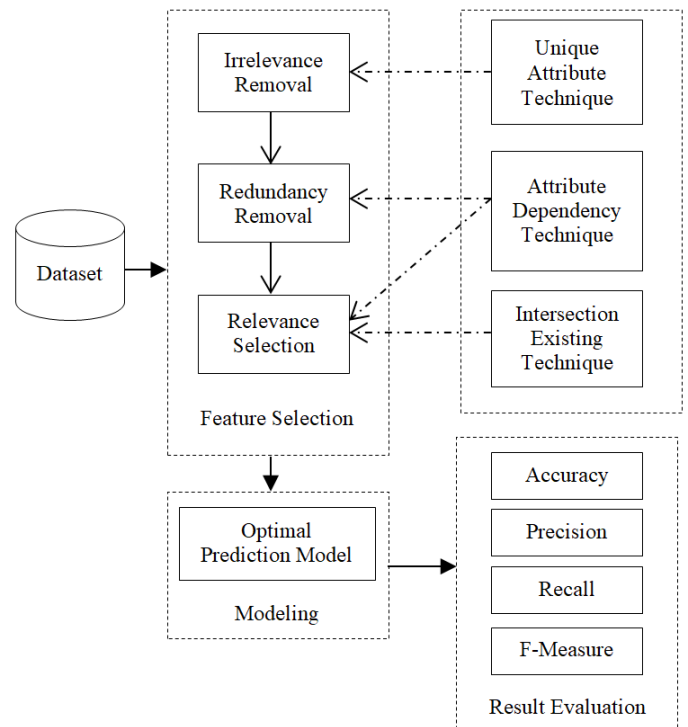


Fig. 1. The proposed framework.

### D. Attribute Dependency to Process the Feature Selection as Proposed Method

Not like the existing technique that selecting features based on statistical measures, attribute dependency selects features based on whether an instance which have the feature value, have the same feature value to others or to the class value. Attribute dependency in the relational model explains the relationship between attributes, or more specifically, the value of an attribute that determines the value of other attributes [24]–[26]. There are several types of dependencies, namely:

1) *Functional Dependency (FD)*: Indicates that if A and B are attributes of relation R, B is functionally dependent on A (denoted  $A \rightarrow B$ ), if each value of A is associated with exactly one value of B.

2) *Full Functional Dependency (FFD)*: Indicates that if A and B are attributes of a relation, B is fully functionally dependent on A if B is functionally dependent on A, but not on any proper subset of A.

3) *Transitive Dependency (TRD)*: A condition where A, B, and C are attributes of a relation such that if  $A \rightarrow B$  and  $B \rightarrow C$ , then C is transitively dependent on A via B (provided that A is not functionally dependent on B or C).

4) *Total Dependency (TD)*: A condition where A and B are attributes of a relation such that if B is functionally dependent on A ( $A \rightarrow B$ ) and A is functionally dependent on B ( $B \rightarrow A$ ).

### E. Proposed Technique

In the proposed work, there are three phases to feature selection. The first phase conducts a preliminary search to remove irrelevant features. In this phase, unique attribute

analysis is employed by identifying features that have a unique value or not, for each row in the dataset. Under this context, features with a unique value means irrelevant and should be removed from the dataset. Another condition is if the features only have one value or if there are more than one value then one value dominates other values, it also means the irrelevant feature, as shown in Algorithm 1 where dataset D as the parameter for configuring the algorithm.

---

Algorithm 1: Irrelevance Removal

---

Input:  
{Dataset} D={F<sub>i</sub>,C}; i=1,2,...,n | F<sub>i</sub> = {f<sub>i1</sub>,f<sub>i2</sub>, .. f<sub>im</sub>} }  
{ where F<sub>i</sub>: features C: class  
n:total number of features  
m:total number of instances  
}  
Result: D<sub>1</sub> {the dataset minus irrelevant features}  
Process:  
i=1;  
D<sub>1</sub> = D  
while i<=n do  
    Get F<sub>i</sub> = {f<sub>i1</sub>,f<sub>i2</sub>, .. f<sub>im</sub>}  
    if (f<sub>i1</sub> ≠ f<sub>i2</sub>...≠ f<sub>im</sub>) or (f<sub>i1</sub> = f<sub>i2</sub>...= f<sub>im</sub>) or  
        (one value dominates other values)  
    then  
        D<sub>1</sub> = D<sub>1</sub> - F<sub>i</sub>;  
    end if  
    i++;  
end  
return D<sub>1</sub>

---

The second phase conducts a search to remove redundant features. Attribute dependency concepts especially total dependency (TD), functional dependency (FD) and full functional dependency (FFD) are employed based on dependency measure in this phase. The first step is applying FFD, the second is applying TD, and the last is applying FD. The advantage of FFD is that it can find not only a relationship between two features but also a group of features (hidden relationship among features).

The hidden relationship among features is the problem that makes the features as redundant features and it cannot be identified by the previous techniques. There is an effort by [31] to address this problem, but only between two features, not for a group of features, and it needs high computational cost. Using FD and FFD, redundant features are identified by measuring the dependency among features or a group of features. The suspected redundant features are easier to find with the help of domain knowledge or domain experts of the dataset. It makes the searching process faster. The more domain knowledge you know the more faster the suspect redundant features you find. By using the FD or FFD evaluation measure, the features or the group of features that are functionally or fully functionally dependent on another feature means that the features or the group of features are redundant and should be eliminated from the dataset. The process of removing the features can be shown in Algorithm 2

below where dataset D<sub>1</sub> as the parameter for configuring the algorithm:

---

Algorithm 2: Redundancy Removal

---

Input :  
{Dataset} D<sub>1</sub>={F<sub>i</sub>,C}; i=1,2,...,n | F<sub>i</sub> = {f<sub>i1</sub>,f<sub>i2</sub>, .. f<sub>im</sub>} }  
{ where F<sub>i</sub>: features C: class  
n:total number of features  
m:total number of instances  
}  
Result: D<sub>2</sub> {the dataset minus redundant features}  
Process:  
D<sub>2</sub> = D<sub>1</sub>  
Repeat  
    { Get the suspected redundant features }  
    { Step 1: utilizing FFD }  
    Get FG = {F<sub>2</sub>, F<sub>3</sub>, ..., F<sub>n</sub>} and F<sub>1</sub> {for example}  
    if FG → F<sub>1</sub> then {F<sub>1</sub> fully FD on FG}  
        D<sub>2</sub> = D<sub>2</sub> - FG  
    end if  
    { Step 2: utilizing TD }  
    Get F<sub>1</sub> and F<sub>2</sub> {for example}  
    if F<sub>1</sub> → F<sub>2</sub> and F<sub>2</sub> → F<sub>1</sub> then {FD each others}  
        D<sub>2</sub> = D<sub>2</sub> - F<sub>1</sub> or D<sub>2</sub> = D<sub>2</sub> - F<sub>2</sub>  
    end if  
    { Step 3: utilizing FD }  
    Get F<sub>1</sub> and F<sub>2</sub> {for example}  
    if F<sub>1</sub> → F<sub>2</sub> then {F<sub>2</sub> FD on F<sub>1</sub>}  
        D<sub>2</sub> = D<sub>2</sub> - F<sub>1</sub>  
    end if  
Until no more suspected redundant features  
return D<sub>2</sub>

---

The third phase conducts a search to select relevant features. The attribute dependency that is used in the second phase is also used in this phase. The difference is only the target that the dependency analysis will be applied between features to the class.

Attribute dependency only yields two values, namely functionally dependent or not, to indicate the dependency between the two or a group of observed features. Related to these, in terms of relevant features, it means there are only two categories, namely relevant features (absolutely strong features) or not relevant features.

Attribute dependency is effective only for absolutely strong features but not for 'not relevant features' (strongly and weakly relevant features). For this reason, in addition to attribute dependency, this study uses a combination of the existing techniques (CHI, IG, GI, Relief, SU) to get the relevant features by applying intersection to the features produced by each technique. The process of selecting the features can be shown in Algorithm 3 below where dataset D<sub>2</sub> as the parameter for configuring the algorithm:

Algorithm 3: Relevance Selection

```

Input :
{Dataset} D2={Fi,C}; i=1,2,...,n | Fi = {fi1,fi2, .. fim} }
{ where Fi: features C: class
  n:total number of features
  m:total number of instances
}
Result: D3 {the selected features}
Process:
D3 = { }
{ Step 1: utilizing FD}
i = 1
while i<=n do
  Get Fi = {fi1,fi2, .. fim}
  if Fi → C then { C FD on Fi }
    D3 = D3 + Fi
  end if
  i++
end
{ Step 2: utilizing the combination of the existing technique}
R, R1 ... R5 = { F1, F2, ..., Fn}
}
D2 = D2 - D3
Get F1 to Fn
R1 = CHI (F,C) {ChiSquared}
R2 = IG (F,C) {Information Gain}
R3 = GR (F,C) {Gain Ratio}
R4 = RF (F,C) {Relief}
R5 = SU (F,C) {Symmetrical Uncertainty}
R = R1 ∩ R2 ∩ R3 ∩ R4 ∩ R5 {Intersection of ALL}

D3 = D3 + R
return D3

```

III. METHODOLOGY

This research methodology consists of four stages. They are data preparation, data processing, modeling and performance analysis.

A. Data Preparation

Related to the goal, this study used the high dimensional public dataset from higher education that consist of irrelevant, redundant and relevant features. The dataset was from the students of Middle East College (MEC), Muscat, Oman, studying in a computing specialization from the sixth semester and above. The dataset was collected from a variety of learning approaches based on Information and Communications Technology (ICT), namely, the Student Information System (SIS), the Learning Management System (LMS) called Moodle, and video interactions from the mobile application called “eDify”. The dataset comprises five modules of data from Spring 2017 to Spring 2021 that consists of 326 student records with 37 features and 1 class in total, including the students’ academic information from SIS (which has 23 features), the students’ activities performed on Moodle within and outside the campus (comprising 10 features), the students’ video interactions collected from eDify (consisting of 4 features) and the outcome of the student either having passed

or failed the module (1 class). The dataset was already clean, so it did not need to clean up the data from the possibility of duplicate data, missing data or other disturbances interfering with the process of classification. It means the data are ready for processing in the next step. An explanation of the data is described in Table II.

TABLE II. DATASET DESCRIPTION

Attribute	Type	Description
	Role	
ModuleCode	nominal	Code of the module in which the student has been registered, such as “Module 1”
	feature	
ModuleTitle	nominal	Title of the module in which the student has been registered, such as “Course 2”
	feature	
SessionName	nominal	Shows the session in which the student has been registered, such as “Session-A”
	feature	
RollNumber	nominal	Identification number of the student, such as “21S1234”
	feature	
ApplicantName	nominal	Name of the student
	feature	
ApplicantMobile	discrete	Mobile number of the student
	feature	
CGPAPoint	discrete	Cumulative grade point average of the student, such as “4.0”
	feature	
CGPADesc	nominal	Category of cumulative grade point average of the student, such as “Very Good”
	feature	
Advisor	nominal	Name of the advisor
	feature	
AttemptCount	discrete	The number of attempts in the module, such as “1”
	feature	
AttemptCountCat	nominal	Category of the number of attempts in the module, such as “Low”
	feature	
RemoteStudent	nominal	Either the student is under remote study mode or not, such as “Yes/No”
	feature	
Probation	nominal	Either the student has a backlog of modules to clear, such as “Yes/No”
	feature	
HighRisk	nominal	The high failure rate in a module, such as “Yes/No”
	feature	
TermExceeded	nominal	Progression rate of the student in the degree plan, such as “Yes/No”
	feature	
AtRisk	nominal	Previously failed two or more modules, such as “Yes/No”
	feature	
AtRiskSSC	nominal	Whether the student been registered by the student success center for any educational deficiencies, such as “Yes/No”
	feature	
SpecialNeed	nominal	Whether the student been registered by the student success center for any special needs, such as “Yes/No”
	feature	
OtherModules	discrete	A student registered in any other modules in the current semester, such as “1”
	feature	
OtherModulesCat	nominal	Category of a student registered in any other modules in the current semester, such as “High”
	feature	
PrerequisiteModules	nominal	Prerequisite module registration, such as “Yes/No”
	feature	

Attribute	Type	Description
	Role	
PlagiarismHistory	discrete	The number of onto which modules the student has been booked for academic integrity violation, such as “1”
	feature	
PlagiarismHistoryCat	nominal	Category of the number of onto which modules the student has been booked for academic integrity violation, such as “Low”
	feature	
CW1	discrete	Marks obtained by the student in their first coursework, such as “86.5”
	feature	
CW1Cat	nominal	Category of marks obtained by the student in their first coursework, such as “Adequate”
	feature	
CW2	discrete	Marks obtained by the student in their second coursework, such as “86.5”
	feature	
CW2Cat	nominal	Category of marks obtained by the student in their second coursework, such as “Excellent”
	feature	
ESE	discrete	Marks obtained in the end semester examination, such as “86.5”
	feature	
ESECat	nominal	Category of marks obtained in the end semester examination, such as “Good”
	feature	
Online C	discrete	User-performed activities within campus (in minutes), such as “25”
	feature	
Online Ccat	nominal	Category of user-performed activities within campus, such as “Poor”
	feature	
Online O	discrete	User-performed activities outside of campus (in minutes), such as “25”
	feature	
Online Ocat	nominal	Category of user-performed activities outside of campus, such as “Poor”
	feature	
Played	discrete	The number of times the video has been played
	feature	
Paused	discrete	The number of times the video has been paused
	feature	
Likes	discrete	The number of times the student has liked the video
	feature	
Segment	discrete	The number of times a student has played a specific portion of the video by using the slider
	feature	
Result	nominal	the outcome of the student either having passed or failed the module
	class	

### B. Data Processing

This phase begins with the identification of the attributes as features and as the class used to build the framework. Based on dataset in the data preparation phase, this study has determined the attribute with label ‘result’ as the class and the rest of attributes as the features as shown in Table II. This process focuses on selecting relevant or strongly features toward the target attribute (class) and removing redundant and irrelevant features or weakly relevant features from the dataset. The prediction model's performance is negatively impacted by redundant and irrelevant features.

Feature selection is one of techniques that greatly impact on enhancing model performance mainly increasing accuracy. In

this research, feature selection is conducted in three stages: (a) irrelevance removal, (b) redundancy removal, and (c) relevance selection.

In the irrelevance removal stage, unique attribute technique is used to test whether the features have (a) a unique value or not, (b) only have one value and (c) one value dominates other values. Features that meet the condition above are removed from the dataset. From this stage, it will get dataset  $D_1$ . The details of the processed are shown in Algorithm 1.

After the irrelevance removal stage, next, it continued to the redundancy removal stage. The total dependency (TD), functional dependency (FD) and full functional dependency (FFD) are used. They are used to test the relationship between two features or features and groups of features. Features or group of features that are functionally (FD) or fully functionally dependent (FFD) on another feature are removed from the dataset. From this stage, it will get dataset  $D_2$ . The details of the processed are shown in Algorithm 2.

After the two stages above, it is followed by relevance selection stage as the last stage. Unlike the two stages before, this stage is employed to select the features used for the construct of the prediction model. There are two concepts used in this stage: (a) attribute dependency, (b) combination of the existing techniques (CHI, IG, GI, Relief, and SU) by applying intersection. Unlike the two stages before, in this stage, the relationship is tested between the features and the class (target attribute). From this stage, it will get the final dataset  $D_3$  as selected features. The details of the processed are shown in Algorithm 3.

### C. Modeling

After obtaining the selected features ( $D_3$ ), the next step is constructing a prediction model. The dataset  $D_3$  is divided into 80% training and 20% testing dataset. Training and testing dataset are used as modeling input. A few of classifiers will be used to process the dataset to classify the student's academic success.

The classifier are Decision Tree (DT), Naive Bayes (NB), Artificial Neural Network (ANN), K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) as the most classifier currently used by researchers to predict student academic success [2]. The process started with training the classifier using the training dataset. The testing dataset tests the classifier in classifying student's academic success. Additionally, depending on measures of classification accuracy and effectiveness, the performance of the classifier model is examined and assessed.

### D. Performance Analysis

This section described how to measure the performance of the prediction model from a number of views since each classifier employed in the modeling process vary. The primary goal of the proposed framework is to achieve the highest accuracy with less number of features. Related to the goal, the confusion matrix is used to determine the accuracy of the prediction model. Comparing the actual value to the predicted value is the approach to assess the model's performance. The confusion matrix is a combination of four different predicted and actual values. In the confusion matrix, the classification

process' outcomes are denoted by four terms: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). According to the confusion matrix, the formula for calculating the accuracy are given in (1)(2)(3)(4) [31].

$$\text{Accuracy} = (TP + TN) / (TP + TN + TP + TN) \quad (1)$$

$$\text{Precision} = (TP) / (TP + FP) \quad (2)$$

$$\text{Recall} = (TP) / (TP + FN) \quad (3)$$

$$\text{F-Measure} = (2 * \text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision}) \quad (4)$$

#### IV. RESULT AND DISCUSSION

The dataset was prepared previously, next used in data processing stage using the proposed feature selection technique.

##### A. Result

Initially, the dataset was processed by applying the Algorithm 1 to remove the irrelevant features and the result is shown in Table III.

TABLE III. IRRELEVANCE REMOVAL

No	Feature	Unique Attribute Measure
1	RollNumber	unique value
2	ApplicantMobile	unique value
3	ApplicantName	unique value
4	RemoteStudent	one value dominated other value (value No : ≥ 99%, value Yes : ≤ 1%)
5	Probation	one value dominated other value (value No : ≥ 96%, value Yes : ≤ 4%)
6	HighRisk	one value dominated other value (value No : ≥ 93%, value Yes : ≤ 7%)
7	TermExceeded	one value dominated other value (value No : ≥ 98%, value Yes : ≤ 2%)
8	AtRiskSSC	one value dominated other value (value No : ≥ 95%, value Yes : ≤ 5%)
9	SpecialNeed	only one value
10	PlagiarismHistory	one value dominated other value (value 0 : ≥ 82%, value 1 : ≤ 17% value 2 : ≤ 1%)
11	PlagiarismHistoryCat	one value dominated other value (value Low : ≥ 99%, value Medium : ≤ 1%)
12	PrerequisiteModules	one value dominated other value (value No : ≤ 8%, value Yes : ≥ 92%)

D<sub>1</sub> = D – Irrelevant Features

D<sub>1</sub> = ModuleCode, ModuleTitle, SessionName, CGPAPoint, CGPADesc, Advisor, AttemptCount, AttemptCountCat, AtRisk, OtherModules, OtherModulesCat, CW1, CW1Cat, CW2, CW2Cat, ESE, ESECat, Online C, Online CCat, Online O, Online OCat, Played, Paused, Likes, Segment

The results in Table III show that there are 12 features that must be removed from the dataset consisting of 1 feature identified as one value (only one value for all instances), 3 features identified as a unique value (different value for each instances), and 8 features identified as one value dominates other value. After removing the irrelevant features, now the rest of the features are: 25 features as a dataset D<sub>1</sub>. The dataset D<sub>1</sub> still have the possibility of redundant features, as a solution, the next step, redundancy removal was carried out. The dataset D<sub>1</sub> was produced by previous stage, then processed by

applying the Algorithm 2 to remove the redundant features and the result is shown in Table IV.

TABLE IV. REDUNDANCY REMOVAL

No	Feature	Attribute Dependency Measure	Type
1	ModuleTitle	ModuleCode ↔ ModuleTitle	TD
2	CGPAPoint	CGPAPoint → CGPADesc	FD
3	AttemptCount	AttemptCount → AttemptCountCat	FD
4	OtherModules	OtherModules → OtherModulesCat	FD
5	CW1	CW1 → CW1Cat	FD
6	CW2	CW2 → CW2Cat	FD
7	ESE	ESE → ESECat	FD
8	Online C	Online C → Online Ccat	FD
9	Online O	Online O → Online Ocat	FD

D<sub>2</sub> = D<sub>1</sub> – Redundant Features

D<sub>2</sub> = ModuleCode, SessionName, CGPADesc, Advisor, AttemptCountCat, AtRisk, OtherModulesCat, CW1Cat, CW2Cat, ESECat, Online CCat, Online OCat, Played, Paused, Likes, Segment

The results (Table IV) show that there are 9 features that must be removed from the dataset consisting of 1 feature identified as Total Dependency (TD), 8 features identified as a Functional Dependency (FD). No one features identified as Full Functional Dependency (FFD). After removing the redundant features, the rest of the features now there are 16 features as a dataset D<sub>2</sub>. The dataset D<sub>2</sub> need further processed to select the relevant or strongly relevant features. It also still have the possibility of weakly or irrelevant features. The last step, relevance selection was carried out by applying the Algorithm 3 based on dataset D<sub>2</sub>. Using the first step in Algorithm 3, no one features functionally dependent on the target attribute (class). It means no feature is selected from this step. Next step, the process continued using the intersection of the combination of the existing technique, and the result is shown in Table V. According to the result as shown in Table V, each technique (CHI, IG, GI, SU) have exactly the same results where the features Paused and Likes are in the lowest rank with the value 0. These two features are irrelevant to the class, so they are removed from D<sub>2</sub>.

While Relief technique have slightly different result where the lowest rank is the feature Paused. Because the value is greater than zero, the feature is still used for processing. After getting the results (R1, R2, R3, R4, R5) of each technique, by applying intersection to the results, the proposed technique got selected features R (Table V) as the final result. The result R consists of 14 features after removing the two lowest ranks. The result R then adds to the result D<sub>3</sub> from the first step. Because in the first step there are no features selected, then D<sub>3</sub> is equal to R that is used for building the model in the step.

TABLE V. RELEVANCE SELECTION

No	Feature Selection Technique	Selected Features	Name of the Result
1	ChiSquared	Advisor, ESECat, Played, ModuleCode, CW1Cat, Segment, CW2Cat, Online Ccat, OtherModulesCat, CGPADesc, Online Ocat, AttemptCountCat, AtRisk, SessionName	R1
2	InfoGain	Advisor, ESECat, ModuleCode, Played, CW1Cat, CW2Cat, Segment, Online Ccat, OtherModulesCat, CGPADesc, Online Ocat, AttemptCountCat, AtRisk, SessionName	R2
3	GainRatio	ESECat, Played, Segment, ModuleCode, Advisor, CW1Cat, CW2Cat, AttemptCountCat, OtherModulesCat, Online Ccat, CGPADesc, AtRisk, SessionName, Online Ocat	R3
4	Symmetrical Uncertainty	ESECat, Played, Segment, Advisor, ModuleCode, CW1Cat, CW2Cat, AttemptCountCat, OtherModulesCat, Online Ccat, CGPADesc, Online Ocat, AtRisk, SessionName	R4
5	Relief	ESECat, ModuleCode, CW1Cat, CW2Cat, Played, Segment, SessionName, Likes, CGPADesc, Online Ccat, Online Ocat, OtherModulesCat, AttemptCountCat, Advisor, AtRisk, Paused	R5
6	Intersection of All	ModuleCode, SessionName, CGPADesc, Advisor, AttemptCountCat, AtRisk, OtherModulesCat, CW1Cat, CW2Cat, ESECat, Online Ccat, Online Ocat, Played, Segment	$R = R_1 \cap R_2 \cap R_3 \cap R_4 \cap R_5$ $D_3 = R$

Evaluation of the prediction model is used to select the best model raised the highest accuracy. Using a dataset split into training and testing data, the value of each model's cross-validation accuracy as an indicator of model performance is evaluated. Table VI shows the evaluation results based on the cross-validation accuracy and F-measure indicators.

TABLE VI. THE PREDICTION MODEL PERFORMANCE FOR CLASSIFICATION OF STUDENT'S ACADEMIC SUCCESS (%)

Classifier	Accuracy	Precision	Recall	F-Measure
DT	92.64	92.6	92.6	92.2
NB	83.13	83.9	83.1	83.5
ANN	84.62	84.1	85.1	82.1
KNN	85.44	85.9	85.4	85.6
SVM	89.57	89.1	89.6	89.1

The result in Table VI shows the proposed framework using the proposed filter selection technique can achieve high accuracy scores for each classifier. The highest accuracy score is 92.64% on the DT classifier, followed by SVM (89.57%), ANN (84.62%), KNN (83.44%) and the last is NB (83.13%).

When compared with the existing technique that is frequently used in predicting academic success, the result is shown in Table VIII. The selected features (dataset) produced by each technique that will be used by classifier is presented in Table VII. The features with the score  $\leq 0$  eliminated from dataset because of the features do not have a relationship to target attribute (class).

TABLE VII. THE NUMBER OF SELECTED FEATURES OF FEATURE SELECTION TECHNIQUE

Feature Selection Technique	Features Category	Sub Total Features	Total Selected Features
CHI, IG, GI, SU	Redundant Features : CW2, CW1, ESE, ModuleTitle, AttemptCountCat, OtherModulesCat, Online Ccat, AtRiskSSC, CGPADesc, Online Ocat, AtRisk, SessionName, RemoteStudent	13	28
	Irrelevant Features : RollNumber, ApplicantMobile, ApplicantName, Probation, HighRisk, TermExceeded, PlagiarismHistoryCat, PrerequisiteModules	8	
	Relevant Features : ESECat, Played, Segment, Advisor, ModuleCode, CW1Cat, CW2Cat	7	
RF	Redundant Features: ModuleTitle, OtherModules, Online O, Online C, CW1, CW2, AttemptCount, CGPAPoint	8	31
	Irrelevant Features: ApplicantMobile, RollNumber, PrerequisiteModules, HighRisk, PlagiarismHistory	5	
	Relevant Features: ESECat, ModuleCode, CW1Cat, CW2Cat, ESE, Played, SessionName, Segment, Likes, Online Ocat, AttemptCountCat, OtherModulesCat, Advisor, AtRisk, CGPADesc, Online Ccat, TermExceeded, Paused	18	
CFS	Redundant Features : CW1, CW2	2	3
	Irrelevant Features: ApplicantName	1	
	Relevant Features: -	0	
Proposed Technique	Redundant Features : -	0	14



Feature Selection Technique	Features Category	Sub Total Features	Total Selected Features
	Irrelevant Features: -	0	
	Relevant Features: ModuleCode, SessionName, CGPADesc, Advisor, AttemptCountCat, AtRisk, OtherModulesCat, CW1Cat1, CW1Cat2, ESECat, Online Ccat, Online Ocat, Played, Segment	14	

The result in Table VII shows the existing techniques successfully reduced the number of features that have not significant affecting to the target attribute. However, the selected features produced by each existing technique still have irrelevant and redundant features, impacting the accuracy of the prediction model is not optimal. The techniques (CHI, IG, GI, SU) have the same result in a number of selected features (28 features) and the features that have been selected. The difference is only the ranking of the selected features. The CFS technique slightly has a different result in producing selected features (only three features). It is an excellent factor in reducing the computational cost. According to these results, it concludes that selected features that still contain redundant and irrelevant features, making the accuracy result is not optimal.

TABLE VIII. THE ACCURACY COMPARISON OF FEATURE SELECTION TECHNIQUE FOR PREDICTING ACADEMIC SUCCESS (%)

Feature Selection Technique	DT	NB	ANN	KNN	SVM
All (No FS Technique used all dataset)	85.88	71.47	81.54	82.21	86.50
CHI	85.89	73.31	82.82	84.96	86.81
IG	85.89	73.31	82.82	84.96	86.81
GI	85.89	73.31	82.82	84.96	86.81
SU	85.89	73.31	82.82	84.96	86.81
RF	85.89	73.31	81.53	82.51	88.03
CFS	84.62	80.98	61.04	81.60	80.98
Proposed Technique	92.64	83.13	84.62	85.44	89.57

The result in Table VIII shows the proposed framework using the proposed filter selection technique compared to existing technique can achieve high accuracy rates for each classifier. According to Table VIII, it can be seen the accuracy of the existing technique is less than the proposed technique due to the existing the irrelevant and redundant features. Based on these results, the proposed technique outperforms the existing technique (CHI, IG, GI, SU, RF), even with the CFS (the filter-based feature subset techniques). The RF technique produced the most selected features, but the accuracy is not quite different from others (CHI, IG, GI, SU as shown in

Table VIII). The RF technique raised the highest accuracy in classifier SVM and the lowest accuracy in ANN and KNN, while in DT or NB have the same result from others (CHI, IG, GI, and SU).

### B. Discussion

According to the results (Table VI), it concludes that the best classifier to classify student's academic success is DT. This statement is supported by the previous studies that appropriate classifier is one of main factor besides feature selection in predicting academic success, and DT leads to the highest accuracy and performs well in a large dataset [1], [2]. This happen because the feature CGPADesc is used as the main feature. The results proved that the proposed framework is feasible to implement. According to Tables VII and VIII, it can be concluded that the criteria of good feature selection are not only determined by the number of selected features but also by the kind of selected features. Even though the result of CFS is the least (Table VII), but the accuracy is the lowest result (as shown in Table VIII). This happen because the result still consist of redundant and irrelevant features and have no relevant features, so it affects to the accuracy of the model [5]. The techniques (CHI, IG, GI, SU) have exactly the same results in each classifier because the dataset (the selected features produced by each technique) is the same. The proposed technique raised the highest accuracy rate even though the number of selected features (14) is greater than CFS (3) due to the nonexistence of the redundant and irrelevant features. It proved that feature selection is another main factor besides the classifier in predicting academic success aligned with the previous studies [1], [2], and the proposed feature selection leads to the highest accuracy and performs well in a large dataset. From the whole result, this research has proven that the framework develop using the proposed feature selection technique reached the highest accuracy score over all the existing technique. This research also proved that the feature selection can increase the prediction model accuracy, and it aligned with the previous studies [17]–[22].

### V. CONCLUSION

Generating graduates with better academic performance through a prediction model is a crucial factor and challenging task in higher education. This study evaluates filter-based feature ranking techniques for predicting academic success and proposes a novel feature selection technique to improve the performance of the prediction model through a framework. The proposed framework will act as a guide that gives suggestions for creating the prediction model. Based on the accuracy rates used to evaluate performance, the best model was chosen. According to the results, it conclude that the best classifier to classify student's academic success is the DT with accuracy = 92.64%. The results show that the proposed framework utilizing the proposed technique significantly improves the accuracy of the proposed prediction models. The proposed technique increase the accuracy compared to existing technique from 2-8%. These results indicate the proposed framework utilizing the proposed technique effectively used to predict student success for the higher education institution making policies or giving intervention to poor academic performance students.

REFERENCES

- [1] A. Hellas *et al.*, "Predicting academic performance: A systematic literature review," *Annu. Conf. Innov. Technol. Comput. Sci. Educ. ITiCSE*, pp. 175–199, 2018, doi: 10.1145/3293881.3295783.
- [2] E. Alyahyan and D. Düşteğör, "Predicting academic success in higher education: literature review and best practices," *Int. J. Educ. Technol. High. Educ.*, vol. 17, no. 1, 2020, doi: 10.1186/s41239-020-0177-7.
- [3] Dafid and Ermatita, "Filter-Based Feature Selection Method for Predicting Students' Academic Performance," *2022 Int. Conf. Data Sci. Its Appl. ICoDSA 2022*, pp. 309–314, 2022, doi: 10.1109/ICoDSA55874.2022.9862883.
- [4] S. García, J. Luengo, and F. Herrera, *Data Preprocessing in Data Mining*, vol. 72. Cham: Springer International Publishing, 2015.
- [5] G. Manikandan and S. Abirami, "An efficient feature selection framework based on information theory for high dimensional data," *Appl. Soft Comput.*, vol. 111, 2021.
- [6] K. P. Shroff and H. H. Maheta, "A comparative study of various feature selection techniques in high-dimensional data set to improve classification accuracy," *IEEE Trans. Knowl. Data Eng.*, 2015, doi: 10.1109/ICCC1.2015.7218098.
- [7] N. Rachburee and W. Punlumjeak, "A comparison of feature selection approach between greedy, IG-ratio, Chi-square, and mRMR in educational mining," *Proc. - 2015 7th Int. Conf. Inf. Technol. Electr. Eng. Envisioning Trend Comput. Inf. Eng. ICITEE 2015*, pp. 420–424, 2015, doi: 10.1109/ICITEED.2015.7408983.
- [8] M. Malekipirbazari, V. Aksakalli, W. Shafqat, and A. Eberhard, "Performance comparison of feature selection and extraction methods with random instance selection," *Expert Syst. Appl.*, vol. 179, no. February 2020, p. 115072, 2021, doi: 10.1016/j.eswa.2021.115072.
- [9] C. Anuradha and T. Velmurugan, "Performance Evaluation of Feature Selection Algorithms in Educational Data Mining," *Int. J. Data Min. Tech. Appl.*, vol. 5, no. 2, pp. 131–139, 2016, doi: 10.20894/ijdm.102.005.002.007.
- [10] W. Zheng, "A Comparative Study of Feature Selection Methods," *Int. J. Nat. Lang. Comput.*, vol. 7, no. 5, pp. 01–09, 2018, doi: 10.5121/ijnlc.2018.7501.
- [11] S. Asim, A. Shah, H. M. Shabbir, S. U. Rehman, and M. Waqas, "A Comparative Study of Feature Selection Approaches: 2016-2020," *Int. J. Sci. Eng. Res.*, vol. 11, no. February, pp. 469–478, 2020.
- [12] M. B. de Moraes and A. L. S. Gradvohl, "A comparative study of feature selection methods for binary text streams classification," *Evol. Syst.*, vol. 12, no. 4, pp. 997–1013, 2021, doi: 10.1007/s12530-020-09357-y.
- [13] B. Xue, M. Zhang, W. N. Browne, and X. Yao, "A Survey on Evolutionary Computation Approaches to Feature Selection," *IEEE Trans. Evol. Comput.*, vol. 20, no. 4, pp. 606–626, 2016, doi: 10.1109/TEVC.2015.2504420.
- [14] B. Ghotra, S. McIntosh, and A. E. Hassan, "A large-scale study of the impact of feature selection techniques on defect classification models," *IEEE Int. Work. Conf. Min. Softw. Repos.*, no. May 2017, pp. 146–157, 2017, doi: 10.1109/MSR.2017.18.
- [15] A. R. S. Parmezan, H. D. Lee, N. Spolaôr, and F. C. Wu, "Automatic recommendation of feature selection algorithms based on dataset characteristics," *Expert Syst. Appl.*, vol. 185, no. July 2020, p. 115589, 2021, doi: 10.1016/j.eswa.2021.115589.
- [16] D. A. A. Gnana, "Literature Review on Feature Selection Methods for High-Dimensional Data," vol. 136, no. 1, pp. 9–17, 2016.
- [17] A. U. Khasanah and Harwati, "A Comparative Study to Predict Student's Performance Using Educational Data Mining Techniques," in *IOP Conference Series: Materials Science and Engineering*, 2017, vol. 215, no. 1, doi: 10.1088/1757-899X/215/1/012036.
- [18] S. Hussain, N. A. Dahan, F. M. Ba-Alwib, and N. Ribata, "Educational data mining and analysis of students' academic performance using WEKA," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 9, no. 2, pp. 447–459, 2018, doi: 10.11591/ijeecs.v9.i2.pp447-459.
- [19] N. Priyasadie and S. M. Isa, "Educational Data Mining in Predicting Student Final Grades on Standardized Indonesia Data Pokok Pendidikan Data Set," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 12, pp. 212–216, 2021, doi: 10.14569/IJACSA.2021.0121227.
- [20] M. Zaffar, M. A. Hashmani, K. S. Savita, and S. S. H. Rizvi, "A study of feature selection algorithms for predicting students academic performance," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 5, pp. 541–549, 2018, doi: 10.14569/IJACSA.2018.090569.
- [21] J. D. Febro, "Utilizing feature selection in identifying predicting factors of student retention," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 9, pp. 269–274, 2019, doi: 10.14569/ijacsa.2019.0100934.
- [22] P. Sokkhey and T. Okazaki, "Study on dominant factor for academic performance prediction using feature selection methods," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 8, pp. 492–502, 2020, doi: 10.14569/IJACSA.2020.0110862.
- [23] B. Venkatesh and J. Anuradha, "A Review of Feature Selection and Its Methods," vol. 19, no. 1, pp. 3–26, 2019, doi: 10.2478/cait-2019-0001.
- [24] Thomas M. Connolly and C. E. Begg, *Database Systems A Practical Approach to Design, Implementation, and Management*. Pearson Education, 2015.
- [25] A. Silberschatz, H. F. Korth, and S. Sudarshan, *Database System Concepts*, 6th ed., no. 6. McGraw-Hill, 2010.
- [26] R. Elmasri and S. B. Navathe, *Fundamentals of Database Systems*, 7th ed. Pearson, 2016.
- [27] N. A. Al-thanoon, Z. Yahya, and O. Saber, "Chemometrics and Intelligent Laboratory Systems Feature selection based on a crow search algorithm for big data classification," *Chemom. Intell. Lab. Syst.*, vol. 212, no. September 2020, p. 104288, 2021, doi: 10.1016/j.chemolab.2021.104288.
- [28] S. L. Shiva Darshan and C. D. Jaidhar, "Performance Evaluation of Filter-based Feature Selection Techniques in Classifying Portable Executable Files," in *Procedia Computer Science*, 2018, vol. 125, pp. 346–356, doi: 10.1016/j.procs.2017.12.046.
- [29] B. Nouri-moghaddam, M. Ghazanfari, and M. Fathian, "A novel multi-objective forest optimization algorithm for wrapper feature selection," *Expert Syst. Appl.*, vol. 175, no. December 2020, p. 114737, 2021, doi: 10.1016/j.eswa.2021.114737.
- [30] M. A. Ambusaidi, X. He, P. Nanda, and Z. Tan, "Building an intrusion detection system using a filter-based feature selection algorithm," vol. 9340, pp. 1–13, 2016, doi: 10.1109/TC.2016.2519914.
- [31] C. C. Aggarwal, *Data Mining: The Textbook*, no. April. Springer International Publishing Switzerland, 2016.

# A Performance Analysis of Point CNN and Mask R-CNN for Building Extraction from Multispectral LiDAR Data

Asmaa A. Mandouh<sup>1</sup>, Mahmoud El Nokrashy O. Ali<sup>2</sup>, Mostafa H.A. Mohamed<sup>3</sup>,  
Lamyaa Gamal EL-Deen Taha<sup>4</sup>, Sayed A. Mohamed<sup>5</sup>

National Authority for Remote Sensing and Space Sciences, Cairo, Egypt<sup>1,4,5</sup>  
Faculty of Engineering-AI-Azhar University, Cairo, Egypt<sup>2,3</sup>

**Abstract**—The extraction of buildings from multispectral Light Detection and Ranging (LiDAR) data holds significance in various domains such as urban planning, disaster response, and environmental monitoring. State-of-the-art deep learning models, including Point Convolutional Neural Network (Point CNN) and Mask Region-based Convolutional Neural Network (Mask R-CNN), have effectively addressed this particular task. Data and application characteristics affect model performance. This research compares multispectral LiDAR building extraction models, Point CNN and Mask R-CNN. Models are tested for accuracy, efficiency, and capacity to handle irregularly spaced point clouds using multispectral LiDAR data. Point CNN extracts buildings from multispectral LiDAR data more accurately and efficiently than Mask R-CNN. CNN-based point cloud feature extraction avoids preprocessing like voxelization, improving accuracy and processing speed over Mask R-CNN. CNNs can handle LiDAR point clouds with variable spacing. Mask R-CNN outperforms Point CNN in some cases. Mask R-CNN uses image-like data instead of point clouds, making it better at detecting and categorizing objects from different angles. The study emphasizes selecting the right deep learning model for building extraction from multispectral LiDAR data. Point CNN or Mask R-CNN for accurate building extraction depends on the application. For building extraction from multispectral LiDAR data, two approaches were compared utilizing precision, recall, and F1 score. The point-CNN model outperformed Mask R-CNN. The point-CNN model had 93.40% precision, 92.34% recall, and 92.72% F1 score. Mask R-CNN has moderate precision, recall, and F1.

**Keywords**—Multispectral LiDAR; Mask R-CNN; Point CNN; deep learning; building extraction

## I. INTRODUCTION

The escalating urbanization of the global population necessitates the development of accurate and efficient techniques for extracting buildings from remote sensing data. The extraction of buildings from remotely sensed data is a crucial procedure with wide-ranging applications, including but not limited to three-dimensional (3D) building modeling, urban planning, disaster assessment, and the maintenance of digital maps and Geographic Information System (GIS) databases [1].

The task of accurately and efficiently identifying buildings from remote sensing data presents several challenges, due to data availability issues, poor data quality, and obstructions caused by nearby objects like trees, automobiles, and

mountains [2]. Despite these difficulties, advancement has been made significant in the recent development of building extraction techniques. Building extraction accuracy and effectiveness are projected to increase over time as deep learning algorithms advance and more high-quality remote sensing data become accessible [3].

The multi-spectral Light Detection and Ranging (LiDAR) provides a field for obtaining different spectral responses from different features and collecting various data about the surface and terrain of the land and water [4]. Due to this rationale, the utilization of multi-spectral LiDAR has significantly advanced the field of remote sensing data due to its vast quantity of high-resolution multispectral and spatial data [5]. However, this abundance of data may present a challenge in terms of the human capacity to accurately extract and classify features from the point cloud. Consequently, the rapid development of computer technology and the emergence of artificial intelligence, including machine learning and deep learning, have made it possible to reduce the time and human effort required for precise feature extraction from LiDAR sensors' point clouds [6, 7]. Automatic extraction of buildings from multispectral LiDAR data is a challenging task, but one that has the potential to be extremely useful for a wide range of applications [7].

The objective of this study is to compare and contrast two distinct methodologies for extracting buildings from multispectral LiDAR data. The first utilizes the deep learning algorithm Mask R-CNN, while the second utilizes Point CNN. Both methods utilize three multispectral LiDAR channels to optimize building extraction.

The paper conforms to this structure. Section II describes the related works. Section III describes the significance of the research. The data and the study area are in Section IV. The Section V describes the methodology. Section VI discusses accuracy assessment mathematically. Section VII provides qualitative and quantitative evaluations of the findings. The discussion and summary concluded in Section VIII.

## II. RELATED WORK

The utilization of LiDAR technology offers a significant advantage in terms of three-dimensional spatial accuracy [8], rendering it an optimal choice for various remote sensing applications, particularly in the mapping of densely populated

urban regions. Multiple scientific studies have provided evidence supporting the utilization of LiDAR data for the extraction of buildings in urban environments [9, 10].

Building extraction from LiDAR data has been extensively researched, resulting in the development of numerous algorithms in recent years. Nevertheless, the majority of these techniques rely on LiDAR data with a single wavelength. There exists a limited body of research pertaining to the utilization of multispectral LiDAR data for the purpose of building extraction. These studies demonstrate that deep learning can be used to accurately extract buildings from single-wavelength LiDAR data. It is essential to note, however, that the efficacy of deep learning models for building extraction can vary depending on the scene's complexity and the quality of the LiDAR data.

One of the earliest studies on building extraction using multispectral LiDAR data was conducted by [11]. They proposed a Graph Geometric Moments Convolutional Neural Network (GGMCNN) model for extracting buildings from airborne multi-spectral LiDAR point clouds. The GGMCNN model is a deep learning model that is specifically designed for processing point cloud data. It takes as input a set of features that are extracted from the point cloud, including the point's elevation, intensity, and spectral information. The GGMCNN model is trained to classify each point cloud as either building or non-building. This study has shown that deep learning can be used to extract buildings from multispectral LiDAR data with high accuracy. However, research is needed to develop and evaluate different deep learning models that deal with point clouds and raster format for this task on more accurate datasets such as multispectral LiDAR data.

Several studies have employed deep learning models to extract buildings from mono-wavelength LiDAR data in a raster format, as evidenced by the works of [3, 12-15]. In contrast, some researchers have studied the extraction of buildings from LiDAR data in the form of point clouds [11]. Nonetheless, a lack of research persists regarding the evaluation of deep learning models that are appropriate for building extraction from multispectral LiDAR data. This is an important area of research, and we hope our study can help fill this gap.

The advent of various deep learning networks specifically designed for processing raw LiDAR data [16-18] has facilitated the direct extraction of buildings from LiDAR point clouds. This is in contrast to previous methods that necessitated data rasterization prior to extracting features from the raster format [19]. Convolutional Neural Networks (CNNs) and their respective lineages are extensively employed and favored networks within the domain of deep learning [20]. One notable advantage of CNNs over previous models is their capacity to autonomously detect significant features without human intervention, rendering them more pragmatic [21]. The deep learning algorithm known as Point Convolutional Neural Network (Point CNN) [19] is distinguished by its capability to directly process raw cloud points, eliminating the requirement for the rasterization process. In a study conducted by [22], a comparative analysis was performed to evaluate the performance of the deep learning algorithm Point CNN in

classifying land points in agricultural areas. The findings of the study demonstrated that Point CNN outperformed traditional methods in terms of accuracy. Furthermore, it has been demonstrated that Mask Region-based Convolutional Neural Network (Mask RCNN), a deep learning algorithm belonging to the same category as CNN, has exhibited notable efficacy in the extraction of buildings from LiDAR data. This is achieved through the conversion of cloud points into raster images [23, 24].

However, to the best of our knowledge, there is a lack of scientific investigations on the automatic extraction of buildings using multispectral LiDAR points based on the Point CNN algorithm, and comparing its results with the Mask RCNN algorithm. Previous studies have not adequately compared these two methodologies. Currently, there exists a significant need for the automated and precise categorization of multispectral LiDAR points across various applications. Also, applying this approach to multispectral LiDAR data is an important step to increase the accuracy of building extraction in complex urban environments and facilitate the emergence of novel applications in subsequent endeavors.

In this study, the main contributions are:

- 1) Compare and contrast two distinct methodologies for extracting buildings from multispectral LiDAR data: Mask RCNN and Point CNN. Both methods utilize three multispectral LiDAR channels to optimize building extraction.
- 2) Investigate the importance of using multispectral LiDAR data for building extraction. Using multispectral LiDAR data can considerably improve the accuracy of building extraction compared to using single-wavelength LiDAR data, according to the study's findings.

Overall, the research on building extraction using deep learning with multispectral LiDAR data is still in its early stages. However, the results from existing studies are promising and suggest that deep learning has the potential to improve the accuracy and efficiency of building extraction in urban environments.

### III. RESEARCH SIGNIFICANCE

This paper presents a significant analysis of two distinct methodologies employed for building extraction from multispectral LiDAR data. The methodologies involve either utilizing point clouds directly or converting them into a raster format. This study demonstrates the effectiveness of deep learning algorithms for building extraction. Furthermore, offers significant insights regarding the utilization of multispectral LiDAR data in the context of building extraction. This study's findings may have a wide variety of practical applications. Urban planners and emergency administrators could use them to automatically generate building footprints, for instance. They could also be used to improve the accuracy of 3D city models.

### IV. STUDY AREA AND DATASET

The study area represents a complimentary dataset provided by the National Center for Airborne Laser Mapping (NCALM), encompassing an urban region located in Houston

in the southeastern sector of Texas, United States of America. The study area, as illustrated in Fig. 1, encompasses the vicinity of the Houston University campus and its immediate surroundings. The study area encompasses an estimated area of 550m<sup>2</sup>. The study area was selected based on its diverse range of building types, encompassing regular residential structures, edifices surrounded by a canopy, and small-scale constructions with haphazard layouts.



Fig. 1. Study area.

The research area was scanned in February 2017 using a Teledyne Optech Titan multi-spectral Airborne Laser Scanning (ALS) system [25]. The ALS dataset included multispectral data from 14 flight lines. All collected points were separately recorded in 42 LAS files, each corresponding to a distinct channel and strip. Every LAS file contained data on the point source ID, scan angle rank, flight line edge, scan direction flag, returns, GPS time, and intensity values. Detailed information about the dataset is shown in Table I; the Titan sensor was put in an Optech aircraft. The flight plan and equipment parameters are shown in the Table II.

TABLE I. SPECIFICATIONS OF OPTECH TITAN MULTI-SPECTRAL ALS LIDAR (TELEDYNE OPTECH TITAN, 2015)

Parameters	Channel 1	Channel2	Channel 3
Wavelength	1550 nm MIR	1064 nm NIR	532 nm Green
Beam divergence	0.35 mrad(1/e)	0.35 mrad(1/e)	0.70 mrad(1/e)
Look angle	3.5 ° forward	nadir	7.0 ° forward
Effective PRF	50–300 kHz	50–300 kHz	50–300 kHz

TABLE II. THE FLIGHT PLAN AND EQUIPMENT PARAMETERS

Flight Parameter	
Sensor ID	The Optech Titan MW (14SEN/CON340) LiDAR sensor
flying height	460 m AGL
swath width	445 m
overlap	50%
line spacing	225 m
Equipment Parameters	
PRF	175 kHz per channel (525 kHz total)
scan frequency	25 Hz
scan angle	±26° and ±2°cut-off at processing

## V. METHODOLOGY

Two strategies for building extraction from multispectral LiDAR data were employed to accomplish the research objective. The first method extracts buildings from the raster data, whereas the second method extracts buildings from the point data.

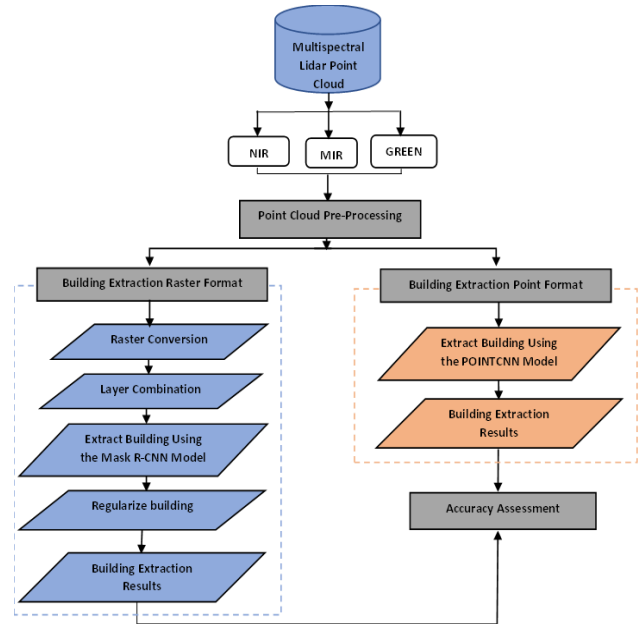


Fig. 2. Pre-processing flow Chart.

The data obtained from Multispectral LiDAR is presented in the LAS format, comprising point cloud values for x, y, z, and intensity. Fig. 2 depicts the flow chart for two distinct approaches to extracting buildings from multispectral LiDAR point cloud data. The building extraction strategy from the raster format is represented in blue, while the building extraction strategy from the point format is represented in orange. The point cloud pre-processing step was executed in two distinct approaches, followed by their respective application to two distinct deep learning algorithms. One algorithm is designed to handle data in raster format, while the other algorithm is specifically designed to process raw data in a point cloud format.

### A. Pre-Processing Point Cloud

The pre-processing of the point cloud is the same for the two methods and it consists of removing the noise points, integrating the point clouds of the three different channels, segmenting the point cloud according to the boundary of the study area, and finally separating the ground and off-ground point clouds. The non-terrain point cloud is of interest in this paper because it contains the features of the buildings. The following will be elaborated upon in a comprehensive manner below:

1) *Data cleaning*: A statistical out-linear removal (SOR) algorithm was utilized to eliminate isolated points or any points that fell outside the intensity range [26]. The logarithmic function operates by computing the spatial separation between a given point and its six adjacent

neighbors. In cases where the mean distance between a given point and its corresponding exceeds the established minimum threshold, the point is eliminated.

2) *Merging and segmenting points*: The absence of control points or reference points in our case has compromised the statistical reporting of geometric quality. In order to evaluate our building extraction methodology, a study area was chosen from a single strip to address the problem at hand. Before performing the step of merging the three different wavelengths into a single LAS file, each channel was cut in three directions based on the borders of the chosen study area. In order to obtain the highest efficiency of the multi-spectral LiDAR points, a merger of the three channels was made, as each channel collects data from a different angle of view and with different intensities, in addition to improving the density of the cloud points. The present study involves the integration of three separate point clouds through the utilization of 3-D spatial join methodology. Each individual point cloud of a specific wavelength serves as the reference point among the three. The reference point cloud employs the closest neighbor searching algorithm to identify neighboring points within the other two wavelengths of point clouds. Subsequently, the segmentation algorithm, which is available via the Cloud Compare software, was employed to clip the multispectral LiDAR points according to the selected boundaries of the chosen study area.

3) *Filtering points*: This research uses deep learning models to extract buildings from multispectral LiDAR data. To simplify this procedure, the Cloth Simulation Filter (CSF) provided by [15] was used to separate ground points and non-ground points as shown in Fig. 3, and concentrate exclusively on the latter because they include buildings. The CSF filter operates through the inversion of the point cloud and drops a simulated cloth model onto the designated points. The cloth undergoes a settling process as it contends with the opposing forces of gravity and internal cloth tension, which occurs over numerous iterations. The filter necessitates the provision of four input parameters, with the first parameter denoting the terrain type. In this case, the area being examined is characterized by flat terrain. Subsequently, a cloth resolution of 2.0, regulates the texture coarseness or smoothness of the cloth's simulation. The terrain simulation's maximum iteration is 500. Ultimately, a classification threshold of 0.5 was established in order to differentiate between terrestrial and non-terrestrial point clouds, utilizing point distances as the determining factor.

### B. Building Extraction of Raster Data

This method utilizes Mask R-CNN and polygon regularization to accomplish its building extraction goals. Mask R-CNN can produce preliminary building polygons from an input image. Then, the basic polygons are transformed into regularized polygons using the polygon regularization technique. The pre-processing stage of the 3D point cloud was conducted to enable the point cloud to be suitable for direct extraction of buildings using the deep learning model Point-

CNN. On the other hand, the deep learning model Mask R-CNN requires several steps to process the multi-spectral LiDAR data, with the most crucial of these steps being the conversion of the three-dimensional point cloud into a horizontal image plane through the process of rasterization.

1) *Raster conversation*: The 3D point cloud was rasterized to a 2D raster with intensity and height information preserved. Cloud intensities at three distinct wavelengths were converted into three distinct intensity images for each channel separately as shown in Fig. 4(a, b, c). The intensity cloud rasterization process was executed by defining a specific set of properties. The cell size parameter was established as 0.1, and the binning interpolation technique was employed to compute the mean intensity value within cells that lack data points. The height raster as shown in Fig. 4(d) was generated from multispectral LiDAR data after merging the three channels and generating a DSM, as opposed to the intensity images for each channel separately. The raster height step was subjected to identical intensity rasterization settings parameters, including interpolation method and cell size, to ensure congruence between the raster resolution and applying the composite process on the four-raster dataset.

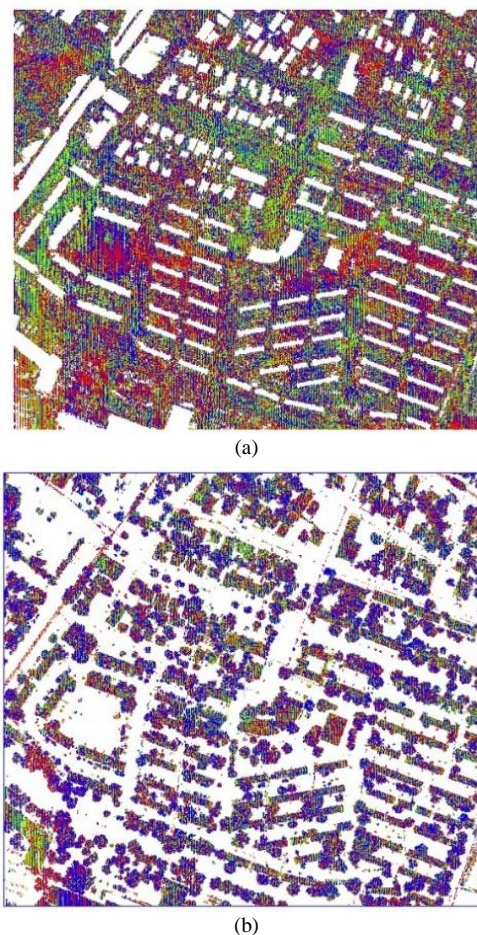


Fig. 3. Separation of ground and off-ground multispectral LiDAR points using CSF algorithm, (a) representing ground points and (b) representing above-ground points.

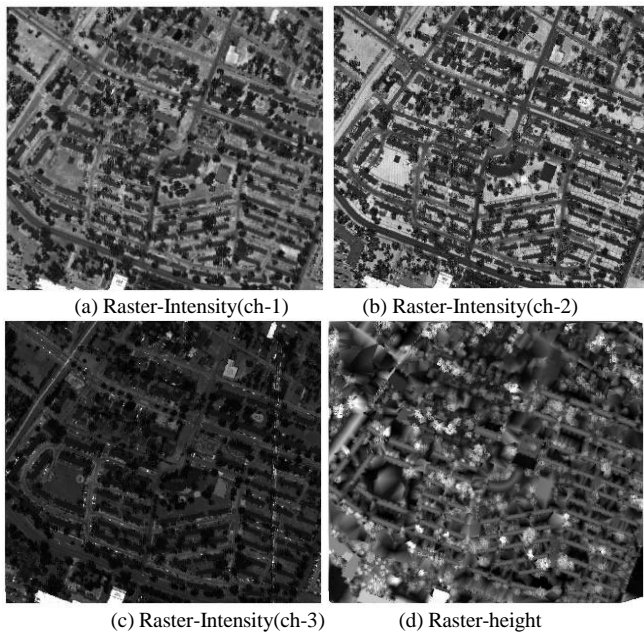


Fig. 4. Rasterization multispectral LiDAR point cloud.

2) *Layer combination*: To streamline the training and validation process of the Mask R-CNN deep learning model, it is necessary to compress the four images within a multi-data image into a new multi-band raster data set. The input dataset comprised a series of four raster images (Raster Intensity Ch-1, Raster Intensity Ch-2, Raster Intensity Ch-3, and Raster height of non-terrain features). All four images possess identical dimensions in terms of length, width, and cell size. The dimensions of the analyzed space were identical to those of the study area, while the depth of the analysis was determined by the quantity of raster images within the input dataset. Specifically, a depth of four was utilized.

3) *Mask R-CNN*: The buildings in this method were extracted using the workflow of the Mask R-CNN deep learning model, as illustrated in Fig. 5. Initially, the multi-spectral LiDAR data underwent a qualification process for its inclusion in the deep learning model, as previously stated [27]. This data is comprised of four images that possess identical length, width, and cell size, and have been compressed accordingly. Subsequently, the training data intended for the deep learning model was exported. To this end, the study area was partitioned into training, validation, and test data sets. Subsequently, the training and validation sets, along with each input image, were utilized to produce the training data set. The input images were partitioned into image tiles of size  $256 \times 256$  pixels, with a 50% overlap step-shift, in order to conform to the input requirements of the Mask R-CNN architecture and to guarantee that all buildings are represented in at least one image tile. The training dataset comprised 184 image slices and 239 features. Following that, the Mask R-CNN model is trained through the utilization of a designated training dataset.

Table III. presents the parameters that were utilized to train the model. The Mask R-CNN architecture comprises several components: including a backbone, a Region Proposal Network (RPN), a Region of an Interest alignment layer (RoI Align), a Bounding box regressor, and a mask generation head [28]. To predict segmentation masks on each Region of Interest (ROI), the Mask R-CNN builds on the Faster R-CNN by adding a network branch to the original (ROI) [24]. To each ROI, the tiny FCN was applied that predicts a pixel-wise segmentation mask for building and nonbuilding regions. The first building polygon is found by following a region's boundary [29]. ArcGIS API for Python, with the Tensorflow and Keras libraries, was used to create and implement the Mask R-CNNs. It's based on a Region Proposal Network (RPN) and ResNet50 backbone.

TABLE III. MASK R-CNN TRAINS THE MODEL PARAMETERS

Parameter	Description
Backbone model	ResNet-50
Training and Test set was split into	70/20%
Validation	10%
Processing type	Nvidia GeForce RTX 2060 GPU
Batch size	4
Learning rate strategy	Stop when the model stop improve
Learning rate	0.0001

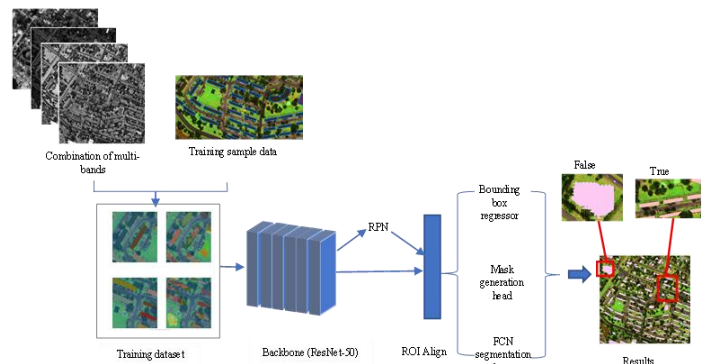


Fig. 5. The Mask R-CNN procedure for building extractions using Multi-Spectral LiDAR Dataset.

4) *Regularize building footprint*: After extracting the buildings with Mask R-CNN, it produces irregular and distorted polygons that don't have straight lines and right angles for edges due to the pixel-labeling location performed by Mask R-CNN [30]. In order to get rid of this randomness in building polygons, the regularized building footprint step was used. A polyline compression algorithm was used to reduce these distortions of building polygons. Table IV shows the parameter used in this algorithm to obtain a much cleaner and closer footprint to buildings than the results that we got from the deep learning algorithm Mask-RCNN.

TABLE IV. REGULARIZE BUILDING FOOTPRINT PARAMETERS

Parameter	Description
Method	Right angle
Tolerance	1
Precision	0.25
Diagonal penalty	1.5
Minimum radius	0.1
Maximum radius	1000000
Processor type	GPU

### C. Building Extraction of Point Cloud Data

The Point Convolutional Neural Network (Point CNN) algorithm is utilized in this research to extract buildings directly from the raw multispectral point cloud without a rasterizing step. The Point CNN architecture consists of an encoder network and a decoder network as shown in Fig. 6, both of which contain X-Conv layers. The cipher network consists of four collective abstraction units to iteratively extract multiscale features of scale (1/256, 1/256, 1/512, 1/1024) concerning the entry point cloud with point number N. Concerning entry points, the entry point merged of intensities of three channels, and adding the elevation point cloud was DSM. K is set from 8 to 16 in this study, where K is the neighboring points around the representative points and N is the number of the point in the previous layer. In this setting, the final point has a 1:0 receptive field since it sees all points from the previous layer, and its features contribute to an accurate semantic interpretation of the shape. In the second X-Conv layer, a dilation rate of  $D = 2$  was used, and then gradually increased in the third and fourth X-Conv layers to ensure that all remaining representative points see the entire figure and are all suitable for making predictions. In this way, the last X-Conv is more thoroughly trained layers, as more connections are involved in the network. The decoder network comprises three feature propagation units, which gradually restore a robust feature mean representation to produce a high-quality classifier point cloud. The output data size of the encoder is  $N/256$ ,  $N/512$ , and  $N/1024$ . Finally, a fully connected layer is added on top of the last output of the X-Conv layer, followed by a loss, to train the network.

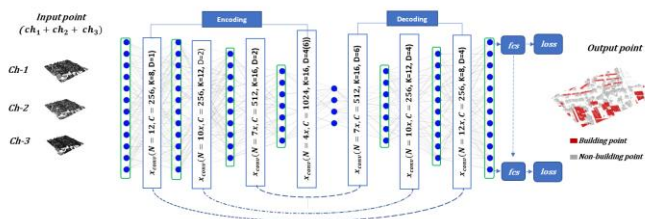


Fig. 6. Point CNN architecture for building extraction from multispectral LiDAR where,  $I_1$ : Intensity of channel-1,  $I_2$ : Intensity of channel-2,  $I_3$ : Intensity of channel-3, N: Number of the point in the previous layer.

## VI. ACCURACY ASSESSMENT

The following measures were employed for assessing the effectiveness of the proposed approaches, all of them are

standard for any semantic segmentation and classification work.

$$Precision = \frac{T_p}{T_p + F_p} * 100\% \quad (1)$$

where the term "precision" refers to the proportion of accurately labeled data points relative to the total number of data points that were labeled, the percentage of data points that were successfully classified relative to the total number of data points that were expected to be classified with this value is the recall. F1-score is the arithmetic mean of the precision and recall values are given as follows:

$$Recall = \frac{T_p}{T_p + F_n} * 100\% \quad (2)$$

$$F1 - score = \frac{2 * precision * recall}{precision + recall} \quad (3)$$

Where:  $T_p$  is a number of point clouds that are classified as true positive building extraction,  $T_n$  are truly negative,  $F_p$  is false positive and  $F_n$  is a false negative.

## VII. RESULTS AND DISCUSSION

The study area was subjected to the training of two deep learning models, namely Mask R-CNN and Point-CNN for the purpose of extracting buildings from a Multispectral LiDAR point cloud. The training dataset and validation for both models were selected using the same buildings to facilitate comparison. The evaluation of building extraction results was conducted using a confusion matrix approach. The reference points for ground truth were obtained from orthorectified aerial photographs captured by the same multispectral LiDAR system in the same survey.

### A. Building Extraction Result of Mask R-CNN

As shown in Fig. 7(a), the outcomes of utilizing the mask R-CNN deep learning model on the multi-spectral LiDAR data after converting them into raster format, as this model confirmed its ability to train and extract the footprints of buildings, but in an irregular style. After they were included in the polyline compression algorithm to create the uniformity of the building, the results are shown in Fig. 7(b).

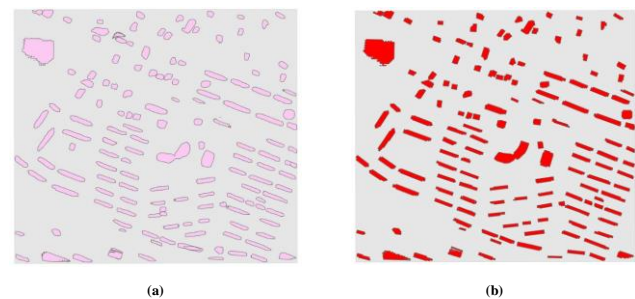


Fig. 7. Building extraction results of mask RCNN: (a) before using regularize, (b) after regularize.

### B. Building Extraction Result of Point CNN

The visual outcomes of building extraction through the utilization of the Point CNN deep learning model are presented in Fig. 8. Specifically, Fig. 8(a) displays the raw multispectral LiDAR points prior to building recognition, with all points



depicted in grey. On the other hand, Fig. 8(b) shows the outcomes attained by the Point CNN model in building recognition, where buildings are highlighted in red and the background is depicted in grey. The employment of Point CNN in contrast to the use of Mask R-CNN is observed to yield superior outcomes. The aforementioned model demonstrated a capacity to discern points of construction with a mean precision of 0.9340 within the given dataset. Furthermore, the algorithm demonstrated a notable proficiency in distinguishing between the points of buildings and those of trees, particularly in cases where buildings were encompassed by thick clusters of trees. This model was also able to extract the points of small and irregular buildings accurately.

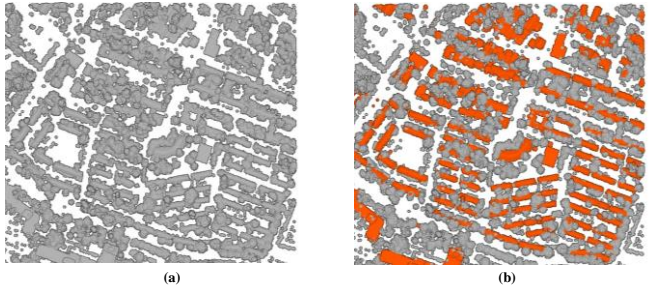


Fig. 8. Building extraction results of Point CNN: (a) before building extraction, (b) after building extraction.

The precision achieved through the utilization of the Point format approach surpasses that of the Raster format methodology. It is anticipated that the inclusion of supplementary spectral data, such as indices, may enhance the ultimate outcomes of building extraction. Ongoing investigations are being conducted with the aim of enhancing the precision of the building extraction.

### C. Comparison between Two Methods

The two methods were compared using evaluation scales such as precision, recall, and F1-score to enhance the comparison. Table V displays the obtained results from using the Point CNN model, which indicates a precision of 93.40%, a recall of 92.34%, and an F1-score of 92.72%. While Mask R-CNN gave less accurate results, it demonstrates a precision of 74.66%, a recall of 67.43%, and an F1-score of 71.17%.

According to the findings of our research, Point CNN performs better than Mask R-CNN when it comes to the extraction of buildings from multispectral LiDAR data in terms of both accuracy and efficiency. In comparison to Mask R-CNN, Point CNN is capable of directly extracting features from point clouds without the need for any pre-processing steps such as voxelization. This results in a greater resolution and significantly faster processing times. In addition to this, Point CNN is able to manage irregular point clouds, which are typical in the case of LiDAR data. Nevertheless, Mask R-CNN continues to have advantages compared with Point CNN in a number of contexts, since it operates on raster data like satellite images or aerial photos rather than point clouds. Also, Mask R-CNN is more suited to identifying and categorizing objects seen from a variety of angles. This makes it an ideal candidate for this task: the Precision, Recall, and F1-score obtained from the two different methods.

TABLE V. THE PRECISION, RECALL, AND F1-SCORE OBTAINED FROM THE TWO DIFFERENT METHODS

Method	Mask R-CNN	Point- CNN
Precision%	74.66	93.40
Recall%	64.43	92.34
F1_Score%	69.17	92.72

TABLE VI. THE TRUE-POSITIVE( $T_P$ ), FALSE-POSITIVE( $F_P$ ), AND FALSE-NEGATIVE( $F_N$ ) FROM THE TWO DIFFERENT METHODS

Method	$T_P$	$F_P$	$F_N$
Mask-RCNN (mask)	112	38	85
Point CNN (points)	897392	26528	91137

According to Tables V, and VI, the results suggest that Point CNN is a more effective method for building extraction from multispectral LiDAR data. It has a higher TP, Precision, and F1-Score than Mask R-CNN. However, Mask R-CNN has a higher Recall, indicating that it is less likely to miss buildings.

Comparing our findings to LiDAR building extraction research [13, 30], our study had 93.40% accuracy, 92.34% recall, and 92.72% F1. Point CNN retrieves features from point clouds without rasterization, which may explain this. Accuracy and processing speed improve. Furthermore, the integration of the three distinct spectra of the multi-spectral LiDAR plays a crucial role in accurately discerning and distinguishing buildings and other features.

According to the study's findings, Point CNN outperformed Mask R-CNN in building extraction from multispectral LiDAR data. This is probable because Point CNN processes point cloud data directly, preserving its structure and characteristics. This enables Point CNN to capture fine-grained geometric details and relationships within the point cloud, which is crucial for accurate building extraction and avoids voxelization, which is sometimes required by Mask R-CNN. This improves efficacy because no data is lost in the conversion process. Maintaining the original spatial distribution of points without voxelization is also essential for accuracy in point clouds with irregular spacing. Due to its architecture, it can manage multispectral LiDAR data from a variety of perspectives. The model captures and uses data from diverse perspectives to increase building extraction accuracy. Mask R-CNN, on the other hand, is well-suited for distinguishing and categorizing objects seen from a variety of angles because it operates on raster data such as satellite images or aerial photographs.

In addition, it is recommended that future research endeavors include evaluating the performance of Point CNN and Mask R-CNN on other datasets, including datasets with different types of scenes (e.g., urban, rural, forested) and datasets with different types of multispectral LiDAR data (e.g., different wavelengths, different point densities, and use of spectral indicators).

## VIII. CONCLUSION

A study was conducted to analyze multispectral LiDAR

data to extract buildings from a residential area situated near the University of Houston, situated in the state of Texas, United States of America. The present investigation undertook a comparative analysis of two discrete deep learning models that are categorized under the Convolutional Neural Network (CNN) family. The present investigation aimed to assess the effectiveness of the method employed in extracting buildings in two separate scenarios. The study involved the utilization of a genuine dataset of multispectral LiDAR data for experimentation purposes. Before inputting LiDAR points into the Point CNN deep learning model, processing operations were executed. Similarly, operations were conducted to transform cloud points into pixels for input into the mask R-CNN deep learning model. Furthermore, a classification of architectural structures was conducted after their acquisition via mask R-CNN. The standardization of ground truth reference was implemented to facilitate a comparison between the two methods, as this is orthorectified aerial photographs captured by the same multispectral LiDAR system in the same survey. It can be concluded that the use of the CNN point model with the proposed approach, which combines the advantages of the intensity of the three different wavelengths plus the height component of the DSM gives better results for extracting buildings from multispectral LiDAR point data, where the accuracy of the results improved by about 30%.

#### ACKNOWLEDGMENT

The authors thank the Hyperspectral Image Analysis Lab at the University of Houston for providing the original Optech Titan data.

#### REFERENCES

- [1] G. Chitturi, "Building Detection in Deformed Satellite Images Using Mask R-CNN," ed, 2020.
- [2] W. Nurkarim and A. W. Wijayanto, "Building footprint extraction and counting on very high-resolution satellite imagery using object detection deep learning framework," *Earth Science Informatics*, vol. 16, no. 1, pp. 515-532, 2023.
- [3] A. Gamal et al., "Automatic LIDAR building segmentation based on DGCNN and Euclidean clustering," *Journal of Big Data*, vol. 7, pp. 1-18, 2020.
- [4] K. Bakula, "Multispectral airborne laser scanning-a new trend in the development of LiDAR technology," *Archiwum Fotogrametrii, Kartografii i Teledetekcji*, vol. 27, 2015.
- [5] O. A. Mahmoud El Nokrashy, L. G. E.-D. Taha, M. H. Mohamed, and A. A. Mandouh, "Generation of digital terrain model from multispectral LiDAR using different ground filtering techniques," *The Egyptian Journal of Remote Sensing and Space Science*, vol. 24, no. 2, pp. 181-189, 2021.
- [6] M. M. Taye, "Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions," *Computers*, vol. 12, no. 5, p. 91, 2023.
- [7] S. S. Ojogbane, S. Mansor, B. Kalantar, Z. B. Khuzaimah, H. Z. M. Shafri, and N. Ueda, "Automated building detection from airborne LiDAR and very high-resolution aerial imagery with deep neural network," *Remote Sensing*, vol. 13, no. 23, p. 4803, 2021.
- [8] A. Novo, N. Fariñas-Álvarez, J. Martínez-Sánchez, H. González-Jorge, and H. Lorenzo, "Automatic processing of aerial LiDAR data to detect vegetation continuity in the surroundings of roads," *Remote Sensing*, vol. 12, no. 10, p. 1677, 2020.
- [9] W. Y. Yan, A. Shaker, and N. El-Ashmawy, "Urban land cover classification using airborne LiDAR data: A review," *Remote Sensing of Environment*, vol. 158, pp. 295-310, 2015.
- [10] I. Prieto, J. L. Izkara, and E. Usobiaga, "The application of lidar data for the solar potential analysis based on the urban 3D model," *Remote Sensing*, vol. 11, no. 20, p. 2348, 2019.
- [11] D. Li et al., "Building extraction from airborne multi-spectral LiDAR point clouds based on graph geometric moments convolutional neural networks," *Remote Sensing*, vol. 12, no. 19, p. 3186, 2020.
- [12] E. Maltezos, A. Doulamis, N. Doulamis, and C. Ioannidis, "Building extraction from LiDAR data applying deep convolutional neural networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 1, pp. 155-159, 2018.
- [13] S. A. Mohamed, A. S. Mahmoud, M. S. Moustafa, A. K. Helmy, and A. H. Nasr, "Building Footprint Extraction in Dense Area from LiDAR Data using Mask R-CNN," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 6, 2022.
- [14] F. H. Nahhas, H. Z. Shafri, M. I. Sameen, B. Pradhan, and S. Mansor, "Deep learning approach for building detection using lidar-orthophoto fusion," *Journal of sensors*, vol. 2018, 2018.
- [15] W. Zhang et al., "An easy-to-use airborne LiDAR data filtering method based on cloth simulation," *Remote sensing*, vol. 8, no. 6, p. 501, 2016.
- [16] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "Pointnet: Deep learning on point sets for 3D classification and segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 652-660.
- [17] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," *Advances in neural information processing systems*, vol. 30, 2017.
- [18] Z. Jing et al., "Multispectral LiDAR point cloud classification using SE-PointNet++," *Remote Sensing*, vol. 13, no. 13, p. 2516, 2021.
- [19] Y. Li, R. Bu, M. Sun, W. Wu, X. Di, and B. Chen, "Pointcnn: Convolution on x-transformed points," *Advances in neural information processing systems*, vol. 31, 2018.
- [20] A. Dhillon and G. K. Verma, "Convolutional neural network: a review of models, methodologies, and applications to object detection," *Progress in Artificial Intelligence*, vol. 9, no. 2, pp. 85-112, 2020.
- [21] L. Alzubaidi et al., "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions," *Journal of Big Data*, vol. 8, pp. 1-74, 2021.
- [22] N. Fareed, J. P. Flores, and A. K. Das, "Analysis of UAS-LiDAR Ground Points Classification in Agricultural Fields Using Traditional Algorithms and PointCNN," *Remote Sensing*, vol. 15, no. 2, p. 483, 2023.
- [23] 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.
- [24] A. Mahmoud, S. Mohamed, R. El-Khoribi, and H. Abdel Salam, "Object detection using adaptive mask RCNN in optical remote sensing images," *Int. J. Intell. Eng. Syst*, vol. 13, no. 1, pp. 65-76, 2020.
- [25] T. O. Titan, "Multispectral LiDAR system: high precision environmental mapping," ed, 2015.
- [26] A. Carrilho, M. Galo, and R. C. Dos Santos, "STATISTICAL OUTLIER DETECTION METHOD FOR AIRBORNE LIDAR DATA," *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, vol. 42, no. 1, 2018.
- [27] K. Yu et al., "Comparison of classical methods and mask R-CNN for automatic tree detection and mapping using UAV imagery," *Remote Sensing*, vol. 14, no. 2, p. 295, 2022.
- [28] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2961-2969.
- [29] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431-3440.
- [30] K. Zhao, J. Kang, J. Jung, and G. Sohn, "Building extraction from satellite images using mask R-CNN with building boundary regularization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 247-251.

# Securing IoT Devices in e-Health using Machine Learning Techniques

Haifa Khaled Alanazi<sup>1</sup>, A. A. Abd El-Aziz<sup>2</sup>, Hedi Hamdi<sup>3</sup>

Department of Computer Science, Jouf University, Al Jouf, Saudi Arabia<sup>1</sup>

Department of Information Systems, College of Computer and Information Science, Jouf University, Aljouf, KSA. Department of Information Systems & Technology, Faculty of Graduate Studies for Statistical Research Cairo University, Cairo, Egypt<sup>2</sup>

Department of Computer Science, Jouf University, Sakaka, KSA. University of Manouba, Tunisia<sup>3</sup>

**Abstract**—The Internet of Things (IoT) has gained significance over the past several years and is currently one of the most important technologies. The capacity to link everyday objects, such as home appliances, medical equipment, autos, and baby monitors, to the internet via embedded devices with a minimum of human interaction has made continuous communication between people, processes, and things feasible. IoT devices have established themselves in many sectors, of which electronic health is considered the most important. The IoT environment deals with many private and sensitive health data that must be kept safe from tampering or theft. If safety precautions are not implemented, these dangers and assaults against IoT devices in the health sector might completely destroy this industry. Detecting security threats to an IoT environment requires sophisticated technology; these attacks can be identified using machine learning (ML) techniques, which can also predict snooping behavior based on unidentified patterns. In this paper, it is proposed to apply five strategies to detect attacks in network traffic based on the NF-ToN-IoT dataset. The classifiers used are Naive Bayes (NB), Random Forest (RF), Decision Tree (DT), Artificial Neural Network (ANN), and Support Vector Machine (SVM) models. These algorithms have been used instead of a centralized method to deliver compact security systems for IoT devices. The dataset was pre-processed to eliminate extraneous or missing data, and then a feature engineering approach was used to extract key features. The results obtained by applying each of the listed classifiers to a maximum classification accuracy of 98% achieved by the RF model showed our comparison to other work.

**Keywords**—IoT; ML; DL; attack classification; e-health

## I. INTRODUCTION

A network of physical items, or "things," that have sensors, software, and other technologies built into them that can connect to and exchange data with other systems and devices through the Internet is referred to as the "Internet of Things" (IoT) [1]. These devices range from basic household goods to cutting-edge industrial equipment. More than 7 billion IoT devices are currently online. IoT Analytics experts predict that by 2023, there will be 14.4 billion linked IoT devices, an increase of 18%, and that by 2025, there may be 27 billion connected IoT devices [2].

The Internet of Things has applications in many different industries. It has proven important in a number of different industries, but the healthcare industry has seen it hard. The medical industry has benefited from modern technology and

digital transformation. As mobile medical devices, mobile health applications, and services have helped improve healthcare services, they are expected to rely more and more on IoT technology in the coming years [3].

The use of the IoT in healthcare is constantly evolving. This fundamental change positively affected patient care because it allowed the clinician to provide a more accurate diagnosis and thus achieve better treatment outcomes [4]. The quality and efficiency of medical services have greatly improved due to the integration of IoT elements into medical devices. Today, IoT technology is widespread in hospitals. It has gotten to the point that many doctors, nurses, and healthcare professionals have abandoned paper in favor of tablets and other Wi-Fi-connected devices [3]. With all these changes and developments, the digital transformation of healthcare has created several difficulties affecting patients, healthcare workers, technology innovators, policymakers, and others. Data interoperability is an ongoing difficulty due to the massive amounts of data generated from various systems that store and encode data differently [5]. These concerns, in turn, raise questions about security and privacy. For example, what if medical devices are hacked? Or what if this sensitive patient data is accessed, leaked, or tampered with?

Data and information are among the most important considerations that must be considered when developing and building IOT to avoid any potential risks related to security. The primary concern of network devices is data protection. Security is paramount in the field of IoT because unauthorized access to or interference with IoT equipment, especially when used for major IoT applications, can endanger human life [6]. The IoT environment deals with a huge amount of private and sensitive health data that must be kept safe from tampering or theft. If safety precautions are not implemented, these dangers and assaults against IoT devices in the health sector might completely destroy this industry. These attacks are often carried out to make money, either by selling the stolen data or by holding the victim's data at ransom to release their data.

The main objective of this research is to build and design a suitable model based on machine learning techniques to increase the accuracy of detecting malicious and benign attacks in an IoT environment using the standard NF-ToN-IoT dataset consisting of network traffic attributes based on different protocols to analyze traffic tracking and behavior of networks and identify malicious attacks. Then, using a

preprocessing and feature selection step, the most important features were extracted, and the dimensionality of the dataset was reduced. Supervised classification algorithms were then applied, which included RF, SVM, DT, ANN, and NB classification algorithms that allow the identification of malicious and benign traffic, which helped in developing an effective intrusion detection system (IDS) that can identify a variety of attacks. The proposed model was then evaluated based on the most appropriate metrics. Finally, compare the proposed model with the latest developments in this field. The contributions of our technology are listed below as detailed previously:

- Proposing an appropriate Machine learning (ML) model for Cyber-attack intrusion detection, by using the NF-ToN-IoT dataset and applying classification techniques that include RF, SVM, DT, ANN, and NB algorithms.
- Evaluation of the proposed model based on the most suitable metrics.
- Comparing the proposed model with state-of-the-art in this field.

The remainder of this research is organized as shown below. The literature review on intrusion detection through the application and use of machine learning and deep learning techniques based on different datasets is summarized in Section II. Section III contains a comprehensive explanation of the methodology proposed for this research. The model implementation and evaluation of the machine learning algorithms implemented in this work and the results of each of them are presented in Section IV, while the obtained result and discussion compared to previous studies are presented in Section V. Finally, the conclusion and future work is presented in Sections VI and VII, respectively.

## II. LITERATURE REVIEW

This section will provide an overview of the available studies and research related to the issue of securing the Internet of Things device in the e-Health system.

Zhu et al. [7] recommend using a nonlinear kernel support vector machine (SVM) to build the e-Diag framework, an efficient and privacy-preserving online medical prediagnosis tool, in an IoT-based e-Health environment. When utilizing e-Diag for online prediagnosis services, sensitive personal health data may be processed without privacy disclosure. On the basis of an improved expression for the nonlinear SVM, a powerful and privacy-preserving classification approach is created, using lightweight multiparty random masking and polynomial aggregation techniques. The SVM classifier and data are hence secure. The approach was tested using the PID database of the UCI machine learning repository.

The focus of the study [8] was on how machine learning affected flow-based anomaly detection in SDN. The authors offered two distinct approaches to intrusion detection systems based on deep neural networks (DNN) and machine learning. The NSL-KDD dataset was employed in the first technique, and feature selection based on the RF classifier led to an

accuracy rate of 82%. However, when paired with DNN-based IDS, the second technique had an accuracy of 88%.

The authors of this study [9] compare the performance of ANN and Random Forest models using a dataset created by combining the benign and malicious datasets for detecting the Mirai virus in relation to seven IoT devices. The NBaloT dataset, which contains information on the features infected with the Mirai virus, is used to propose a novel technique that relies primarily on machine learning technology. In order to avoid over-fitting, the data partitioning approach known as mutual validity was applied, and ANN was used to conduct the experiment. The accuracy attained is 92.8%. The Opcode data collection, which includes 69,860 harmful programs and 70,140 examples of common malware, was employed in this study.

The study [10] is built on a deep learning-based technique for Internet of Things intrusion detection. The researchers discovered that security vulnerabilities rise as the number of Internet of Things devices rises. As a result, a Bot-IoT data set was utilized to compare deep learning techniques like CNN with machine learning techniques like RF and MLP. Through their expertise, CNN attained the maximum accuracy of 91% and the lowest accuracy of 88%.

A novel deep learning-based intrusion detection system for IoT networks and devices is presented by the authors in another paper [11]. A four-layer fully connected (FC) deep network architecture is used by this system to identify malicious traffic that may be used to target linked IoT devices. In order to simplify deployment, the suggested system was created as a communication protocol-independent solution. The ground-breaking IDS powered by deep learning maintain an average detection rate of 93.21%. It made use of the DID dataset.

With the use of outside resources, the authors of this study [12] suggest a comprehensive multi-level, privacy-preserving SVM training and illness diagnosis system. For encrypted data, certain efficient fundamental operational algorithms have been developed. Next, a model training procedure that is effective and protects privacy was created utilizing fundamental operational methods. Then, using the BFV coding system and cryptography method, they created a successful Internet-assisted illness diagnostic scheme. The user only needs to execute a limited number of encryption and decryption operations under their suggested method, which uses cloud servers to accomplish the majority of the illness diagnoses. The efficiency of Internet-assisted illness diagnosis has increased by 85.4% with a total computing cost of 0.175 seconds.

In [13] their study Detecting Cyber Intrusion Using Machine Learning Classification Techniques, the authors demonstrate how artificial intelligence, in particular machine learning techniques, may be utilized to develop a workable data-driven intrusion detection system. Numerous well-known machine learning classification algorithms, such as the Decision Tree, Random Decision Forest, Random Tree, Decision Table, and Artificial Neural Network, have been used to detect intrusions as a result of providing intelligent services in the field of cybersecurity. They conducted studies

to achieve an accuracy of 94% in RF using KDD'99 cup datasets.

The authors of [14] proposed an algorithm to predict patients' current health status in addition to continuing professional monitoring. Additional parameters and methods, such as K-nearest neighbor, logistic regression, support vector machines, random forests, and Adaboost classifiers, are considered using the UCI heart disease dataset. They were successful in providing a tool that would aid patients, medical professionals, and the healthcare system. This way of decision-making is 93% accurate.

IoMT systems are described in [20] as having to manage a sizable amount of data that might be utilized for illness diagnosis, prediction, and monitoring. Medical data about patients should be transferred to cloud storage and external computer devices, respectively, as certain IoMT devices have limited storage and processing capabilities. Security and privacy risks may arise as a result of this approach. A swarm neural network-based approach for identifying intruders in IoMT has been offered as a solution to these problems. The suggested model explores the possibility of properly and effectively evaluating healthcare data as well as identifying intruders during data transfer. An NF-ToN-IoT dataset was used to assess the system's performance, and the findings show that the suggested model achieves 89.0% accuracy.

Based on the above, after examining all the previous studies that were collected and explaining their work and results in detail, it becomes clear that all of them achieved different results through the use of different data sets and algorithms, but the highest accuracy result among them was 94%, while the methodology proposed in our work achieved accuracy higher than all previous studies, reaching 98%.

### III. METHODOLOGY

In this section, the choice of the dataset, as well as the description of its specifics, is discussed. In addition, the dataset pre-processing techniques as well as the feature selection are deliberated. Finally, the model implementation is described.

#### A. Dataset Selection and Description

NF-ToN-IoT was chosen as the data set for this research since it contains a variety of heterogeneous data sources collected from Telemetry datasets of IoT sensors. It contains ten features, and nine types of attacks, namely, (XSS, DDoS, DoS, password cracking attacks, reconnaissance, or verification, MITM, ransomware, backdoors, and injection attacks) [15] [16]. The assaults detected in the dataset used in this study may be classified and defined using the terms below:

1) *XSS attack*: Using XSS technology, malicious code can be injected into trusted Internet applications, such as the web pages of Internet of Things services. In XSS assaults, the attacker transmits malicious code to several end users via an online application, typically a browser-side script.

2) *DDOS attack*: Most of the time, a botnet—a collection of compromised machines—conducts this kind of attack. The

victim's IoT resources are flooded and depleted by this attack's many connections.

3) *DoS attack*: A DoS attack is any attempt to compromise the resources and services of an IoT network. Making IoT services inaccessible is the goal of such an attack.

4) *Password cracking attack*: This hacking approach is used to guess potential password combinations until the precise password is found. Examples include dictionary attacks and brute force assaults. Passwords for IoT services, operating systems, and web apps placed on the test bench can be cracked using this technique.

5) *Scanning attack*: This attack seeks to gather details about test bed network victims' computers, such as active IP addresses and open ports. This assault is the initial phase of a penetration test, often known as an investigation or a cyber-death chain model.

6) *MITM attack*: This kind of attack may happen when hackers place themselves in the middle of users and programs to watch over them or appear to be one of them, creating the false impression that information is flowing normally. Data about networks, online apps, and IoT services might be taken in these hacking scenarios.

7) *Ransomware attack*: It is an advanced form of malware assault that encrypts systems or services and renders them inaccessible to regular users until they pay a ransom. IoT devices and applications might be the target of ransomware attacks since they carry out essential functions.

8) *Backdoors attack*: An attacker can use backdoor malware to obtain unauthorized remote access to infected IoT systems. By controlling infected IoT devices, this threat may launch botnet-based DDoS attacks.

9) *Injection attack*: Attackers use injection techniques to introduce real or fake input data from clients into their targets' systems, such as SQL injection to attack ASP and PHP programs.

Compared to other datasets, the NF-ToN-IoT dataset is appropriate for IoT since it captures the heterogeneous character of contemporary IoT networks. Regarding the statistics of the dataset, the NF-ToN-IoT dataset comprises a total of 1,157,994 rows. The number of rows varies for each attack type, with injection attacks having the highest number of rows (460,812) and ransomware attacks having the lowest number of rows (142). Here is a breakdown of the number of rows for each attack type and benign type in Table I:

TABLE I. STATISTICS OF DATASET

Label	Count
Benign	198450
Backdoor	17243
Ddos	197680
DoS	17056
Injection	460812
Mitm	1288
Password	144792
Ransomware	142
Scanning	20618
Xss	99913

The NF-ToN-IoT includes telemetry data from linked devices, Linux operating system data, Windows operating system logs, and IoT network traffic, among other data sources acquired from the entire IoT system. A medium-scale IoT network provides diverse data. The UNSW Canberra IoT Labs and the Cyber Range designed NF-ToN-IoT. Furthermore, the NF-ToN-IoT is represented in CSV format with a labeled column indicating attack or normal and a sub-category attack type. A CSV is a comma-separated value file, which allows data to be saved in a tabular format. CSV files can be used with most any spreadsheet program, such as Microsoft Excel.

### B. Exploratory Data Analysis

Exploratory Data Analysis (EDA) is one of the essential procedures that can be done on a dataset for several reasons. EDA aims to familiarize the user with the data and provide an understanding of how the data is distributed. Additionally, EDA allows the identification of patterns and relationships between parameters present in the data. EDA is also important because it provides insight into the selection of data and aids in the perfect execution of machine learning tasks [17]. This research used different visualizations as an EDA procedure to understand the NF-ToN-IoT dataset.

The bar graph in Fig. 1 illustrates the different categories of attacks that are present in the selected NF-ToN-IoT dataset, where it contains benign attacks, dos, injection, DDoS, scanning, password, Mitm, XSS, backdoor, and ransomware attacks. The bar graph representation of the dataset shows that the dataset is imbalanced, where different counts of the categories can be seen.

The attack with the most count in the NF-ToN-IoT dataset is the injection, where there are more than 400,000 attacks belonging to it. The second most common attacks are DDoS

and benign attacks, where there are approximately 200,000 of each one of them. The other attacks vary in number, whereas the Mitm and the ransomware attacks are absent in this dataset (zero count).

Furthermore, several features within the dataset can be visualized through histograms to show their form of distribution. For instance, Fig. 2 above represents the distribution of TCP\_FLAGS, FLOW\_DURATION, IN\_PKTS, OUT\_PKTS, PROTOCOL, OUT\_BYTES, L4\_DST\_PORT, and L4\_SRC\_PORT. The representations in Fig. 2 show that all of the features have skewed distributions, whereas only L4\_SRC\_PORT and TCP\_FLAGS have a relatively uniform distribution. The skewness of the features means that they are not well-rounded around the mean value of the feature, which could affect the output results. For this reason, the features with skewness might be subjected to transformation in order to be suitable for use in ML techniques.

### C. Model Architecture

The models were trained using the chosen features representing attack behaviors in this phase. The proposed models were then tested by comparing test data to training data to determine the accuracy of each of the models. The model was only considered ready if the accuracy test was satisfactory; otherwise, the model was retrained until an acceptable accuracy was reached. For performance comparison, algorithms like ANN, RF, SVM, DT, and NB classification techniques were implemented.

These algorithms were chosen because researchers widely use them to detect unusual traffic. Furthermore, they are simple and light, they do not require many operations, they are highly accurate, and they have fewer input features. The architecture of our proposed framework is described in Fig. 3.

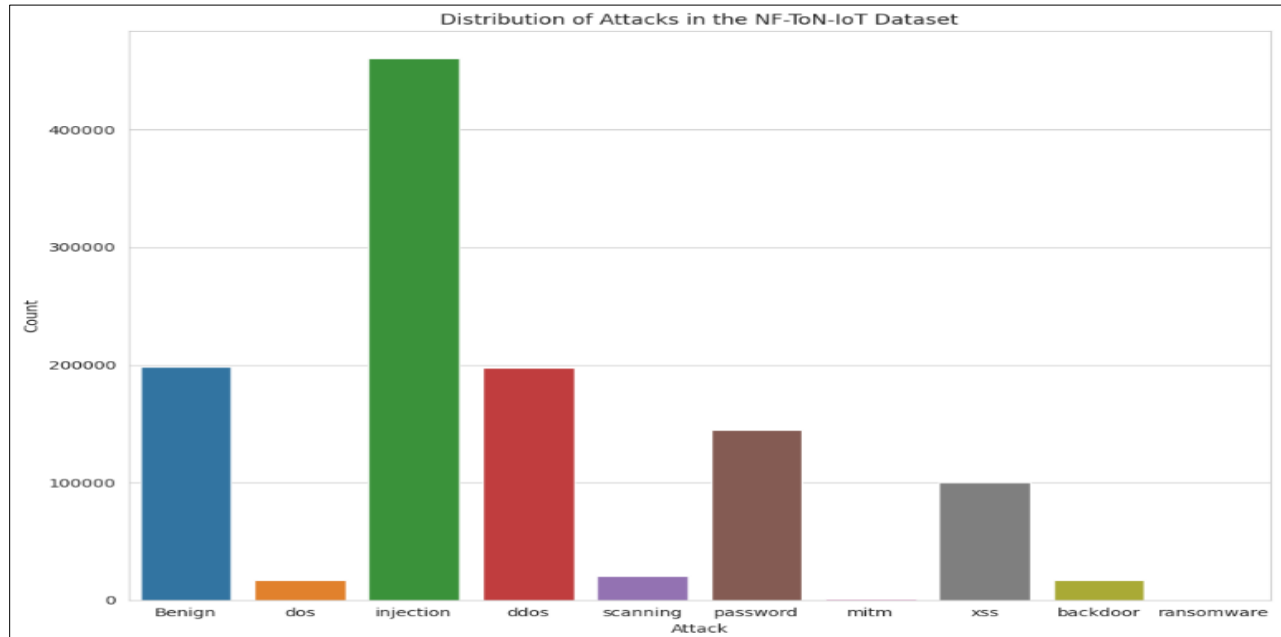


Fig. 1. Distribution of attacks in the NF-ToN-IoT dataset.

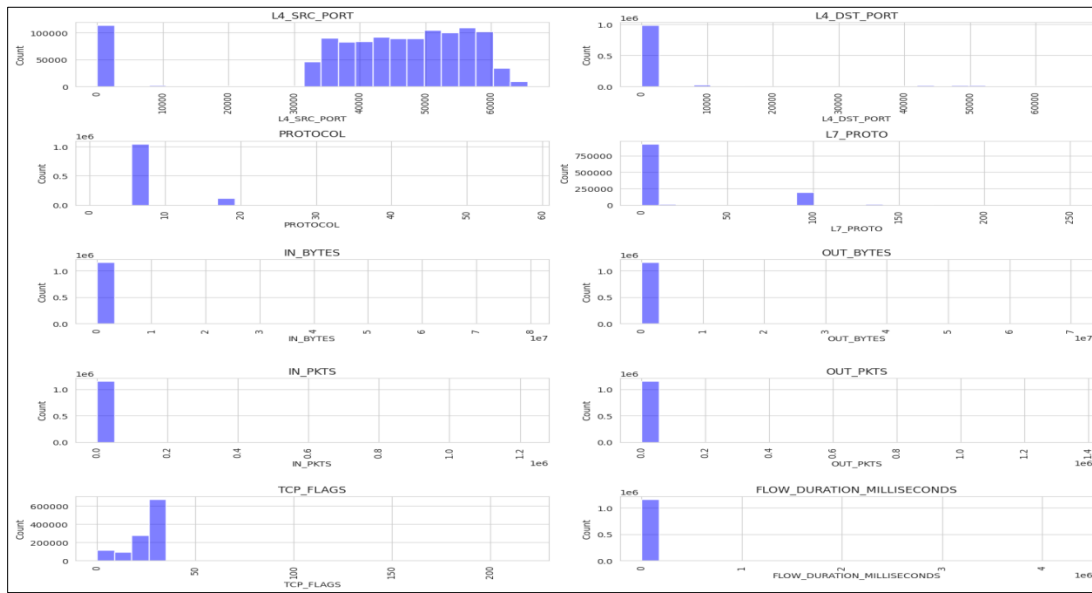


Fig. 2. Distribution of several features in the NF-ToN-IoT dataset.

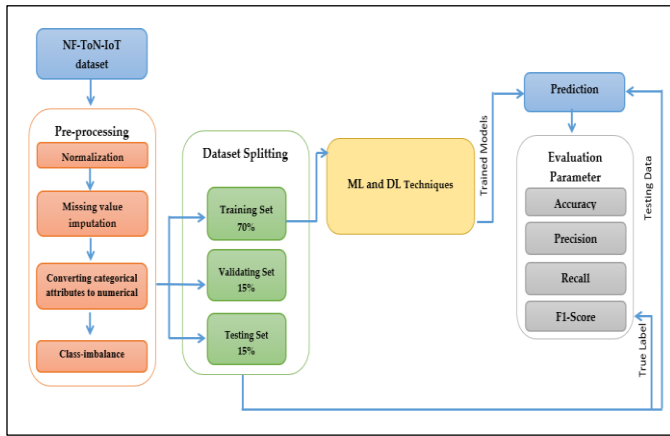


Fig. 3. Architecture of the proposed framework.

As shown in Fig. 3, our framework starts with the acquisition of the NF-ToN-IoT dataset. Then, an essential step is performed, which is the data pre-processing including normalization, imputation of missing values, conversion of categorical attributes to numerical ones, and fixing class imbalance. After the completion of pre-processing, the dataset is subjected to splitting into three different sets: training set (70%), validation set (15%), and testing set (15%). The training set is used for training the ML models (namely, NB, RF, DT, ANN, and SVM). After the training step, the models performed predictions on the testing set in order to find out if the model could accurately generate a result. Different parameters were used for evaluation, such as accuracy, precision, recall, and F1 score.

#### D. Dataset Pre-Processing

Converting unprocessed data into a format that can be read, accessed, and analyzed is known as data pre-processing. Before using ML and DL algorithms, pre-processing is crucial for ensuring or improving any system's overall performance or accuracy. Data is cleaned during the pre-processing phase to

serve a variety of purposes. Some ML algorithms need data in a specific format to ensure the data collection is suitable for many algorithms. The NF-ToN-IoT dataset used in this research presented several challenges, such as missing values, categorical attributes, and class imbalance. To address these issues, the following pre-processing steps were performed:

1) *Missing value imputation*: The NF-ToN-IoT dataset was checked for missing values, but after inspection, it was found that no missing values were found. Therefore, the data set was found to be of high quality and value.

2) *Converting categorical attributes to numerical*: As shown in Table II, the NF-ToN-IoT dataset contains different category features, and the category's attributes must be given numerical values. The conversion process was carried out through LabelEncoder.

TABLE II. CATEGORICAL ATTRIBUTES CONVERTED TO NUMERICAL

Label	Encoded Label	Count
Benign	0	198450
Backdoor	1	17243
Ddos	2	197680
Dos	3	17056
Injection	4	460812
Mitm	5	1288
Password	6	144792
Ransomware	7	142
Scanning	8	20618
Xss	9	99913

3) Class imbalance: Distributions with class imbalances afflict the NF-ToN-IoT dataset. The problem of class imbalance often arises when some classes are far more prevalent than others. Standard classifiers frequently disregard the little classes in these situations because they are too overwhelmed with the large classes.

To address this issue, the SMOTENN (Synthetic Minority Over-sampling Technique and Edited Nearest Neighbors) method was applied to balance the classes in the dataset. The SMOTENN method oversamples the minority class using synthetic data generation (SMOTE) and removes noisy samples using nearest neighbors (ENN) to balance the dataset. Table III shows the number of rows for each attack type before and after applying the SMOTENN method:

TABLE III. DISTRIBUTION OF CLASSES BEFORE AND AFTER SMOTENN

Label	Before	After
Benign	198450	434938
Backdoor	17243	456873
Ddos	197680	200376
Dos	17056	268998
Injection	460812	126022
Mitm	1288	405380
Password	144792	62216
Ransomware	142	457791
Scanning	20618	220336
Xss	99913	114330

### E. Feature Selection

Giving each potential feature a score before choosing the top features is the feature selection procedure. For intrusion detection, many factors must be examined; certain features will be useful, while others will be useless. Each prospective feature is given a score as part of the selection process, which selects the best (k) attributes [18].

A function of both is obtained by independently calculating the frequency of a feature in training for each positive and negative class occurrence. Removing non-essential features improves performance by reducing overfitting, speeding up the calculation, and enhancing accuracy. We'll utilize the filter method Chi2 for the feature selection technique. A statistical method called the Chi2 technique filters out features that aren't as dependent on the class labels as others, and it calculates a score based on feature dependency [19]. The selected features and their scores obtained from the Chi2 technique are presented in Table IV. The scores are calculated based on the frequency of each feature in the training data for both positive and negative class occurrences. The top (k) attributes with the highest scores are chosen for the final feature set.

The selected features contribute significantly to the intrusion detection task by improving the model's performance. Therefore, the seven features shown in Table IV were chosen and were considered the best in terms of influencing the classification process, while if the number of features were increased, they would not have an impact on the classification process. Removing non-essential features reduces overfitting, speeds up computation, and enhances model accuracy.

TABLE IV. SELECTED FEATURES AND THEIR SCORES

Feature	Score	Selected
L4_SRC_PORT	124913.19	Selected
L4_DST_PORT	1021455.86	Selected
PROTOCOL	75773.79	Selected
L7_PROTO	163804.13	Selected
IN_BYTES	104.95	Selected
OUT_BYTES	86.81	Not selected
IN_PKTS	105.52	Selected
OUT_PKTS	28.62	Not selected
TCP_FLAGS	46235.03	Selected

## IV. MODEL IMPLEMENTATION AND EVALUATION

In this section, the obtained results of the proposed models are discussed, starting with exploratory data analysis, followed by an evaluation of each of the proposed models. In this study, the proposed models to be evaluated are Random Forest RF, Support Vector Machine SVM, Decision Tree DT, Artificial Neural Network ANN, and Naïve Bayes NB classifier.

### A. Evaluation Metrics

A variety of methodologies for assessing the effectiveness of the ML techniques employed are chosen and specified in order to offer a thorough and accurate description of the findings achieved. In this paper, the used models were trained on an NF-TON-IoT dataset, and then a collection of data isolated from these models was utilized to assess the trained models' accuracy by correctly separating all the data into its various labels. Accuracy, Precision, F1-Score, and Recall are the measures used to evaluate the algorithms' efficacy. Each equation can be defined separately as follows:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

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

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1 Score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}}$$

1) *Accuracy*: if the classifier can correctly classify the data points, then it is said to have a high accuracy. Accuracy is



represented by the number of correctly predicted instances divided by the total number of predictions.

2) *Precision*: Precision takes into consideration the positive outcomes by the model in comparison to all of the positive outcomes whether they were correctly predicted or not. For this reason, the equation of precision is the ratio of correctly predicted positive outcomes (TP) over all of the positive outcomes (TP and FP).

3) *Recall*: Recall is also known as sensitivity, which represents the fraction of the truly identified positive predictions over the total number of positive instances (represented by TP and FN) because FN should have been predicted as positive results.

4) *F1 Score*: The F1 score combines both precision and recall into a single metric that provides a balanced evaluation of the model's performance. It is a useful metric when there is an uneven class distribution such as in the NF-ToN-IoT dataset.

In all of these equations:

TP stands for True Positive which represents the number of correctly predicted positive instances.

TN stands for True Negative which represents the number of correctly predicted negative instances.

FP stands for False Positive which represents the number of falsely predicted positive instances.

FN stands for False Negative which represents the number of falsely predicted negative instances.

**B. Models' Evaluation**

The evaluation of the different algorithms used is presented separately in this section, which includes RF, SVM, DT, ANN, and NB classifiers. Based on their outcomes on the test dataset, the suggested algorithms' performance is assessed. The performance of models may be assessed using a number of measures, including accuracy, precision, recall, and F1 score. This is how it appears:

- Naïve Bayes:

A supervised machine learning method called the Naive Bayes NB classifier is utilized for classification tasks like text categorization. It also belongs to the family of generative learning algorithms, which implies that it attempts to simulate how an input's distribution varies depending on the class or category.

Upon testing, the Naïve Bayes classifier was able to achieve 72.75% accuracy, which is equivalent to 0.7275. This value indicates that 72.75% of the instances were correctly predicted by the NB classifier. The NB classifier also achieved 0.7567 precision values, which means that out of all of the positively identified instances, 75.67% of them were correctly predicted by the model. A 0.7275 value for recall indicates that the NB model identified 72.75% of the actual positive instances as true positive. Finally, the F1 score achieved by NB was 0.7051 which is a low value, and it means that the overall performance of the model was around 70% well. These

numbers are illustrated in Table V. In addition, a Confusion Matrix Heatmap was generated for the Naive Bayes Classifier in Fig. 4:

TABLE V. NAIVE BAYES CLASSIFIER PERFORMANCE METRICS

Metric	Value
Accuracy	0.7275
Precision	0.7567
Recall	0.7275
F1 Score	0.7051

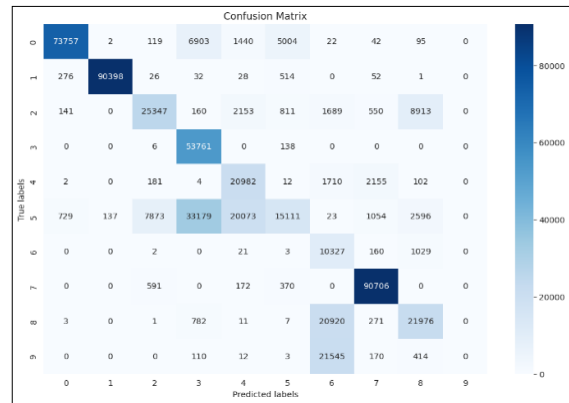


Fig. 4. Confusion matrix heatmap for the naïve bayes classifier.

- Random Forest:

Classifier with Random Forests This classifier is utilized because of its improved accuracy and because it bases its final prediction on predictions from several decision trees rather than just one. Even if the settings are left alone, this supervised machine-learning method produces great results. In order to establish a final categorization of the attack and normal data, tree prediction was also applied.

The random forest model achieved 98.41% accuracy, a 0.9840 precision value, a 0.9841 recall value, and a 0.9840 F1 score, as shown in Table VI. The values shown in the table indicate that the RF classifier achieves a high accuracy rate, where it correctly classified 98.4% of all of the instances in the data. Similarly, it correctly identified the true positives in 98.4% of the total positive instances and the actual positive instances (precision and recall, respectively). Furthermore, the overall performance of the RF model is represented by the high F1 score, which is 0.9840. Fig. 5 shows the Confusion Matrix Heatmap for the Random Forest classifier.

TABLE VI. RANDOM FOREST CLASSIFIER PERFORMANCE METRICS

Metric	Value
Accuracy	0.9841
Precision	0.9840
Recall	0.9841
F1 Score	0.9840

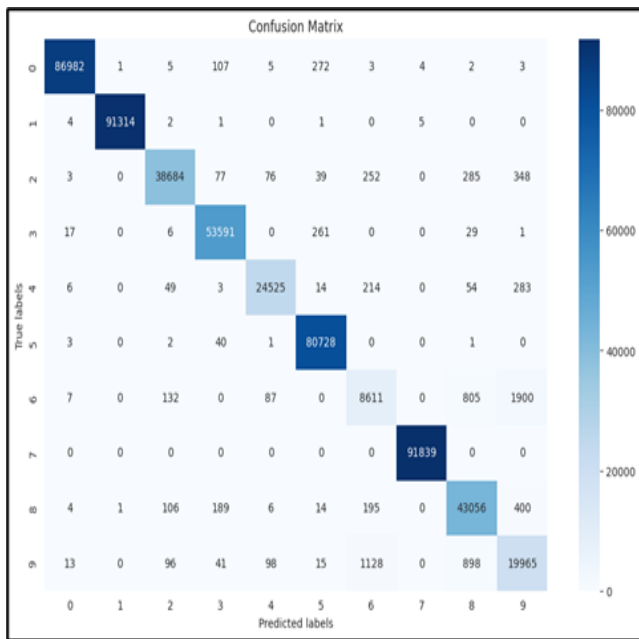


Fig. 5. Confusion matrix heatmap for the random forest classifier.

- Decision Tree:

DT is a well-known construction method that interpolates leaves and branches to resemble a decision tree, where the inner node stands in for the classification rule and the leaves for the class label. The branch also indicates the outcomes. Using the information gained the best branch and root node properties are chosen during the training phase. A decision node is then constructed using the most information gained. As a result, a new sub tree is established beneath the decision node. As the final value will be determined and utilized as the output value, this method will only end if all items in the chosen subgroups have the same value. If there is just one node in the subgroup and no other options, the cycle may also come to an end.

The evaluation metrics for the DT classifier are shown in Table VII and the confusion matrix heat map is in Fig. 6. The Decision Tree classifier accomplished an accuracy of 97.08%, suggesting that the model accurately classified most of the data instances. With a precision score of 97.06%, it showed a high level of correctness when predicting instances as attacks. The recall score of 97.08% indicates the classifier's ability to identify actual attacks accurately from all the positive instances as well. The F1 Score of 97.07% was high, showing a good performance of the DT model.

TABLE VII. DECISION TREE CLASSIFIER PERFORMANCE METRICS

Metric	Value
Accuracy	0.9708
Precision	0.9706
Recall	0.9708
F1 Score	0.9707

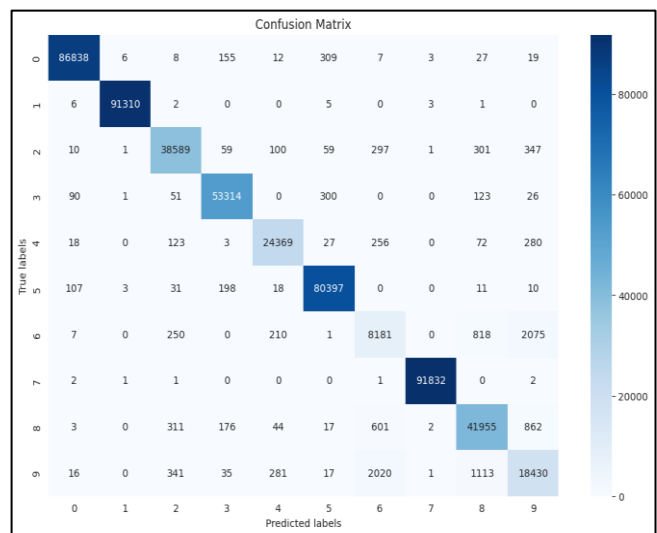


Fig. 6. Confusion matrix heatmap for the decision tree classifier.

- Artificial Neural Network:

An ANN is a collection of linked input-output networks with a weight assigned to each network. As an affiliation, one input layer and one or more intermediary layers make up this structure and only one output layer. The Artificial Neural Network reached an accuracy of 93.20%, demonstrating its ability to classify data instances with a high level of correctness. With a precision score of 93.16%, the classifier presented a strong accuracy when predicting instances as attacks. The recall score which is also the sensitivity value is 93.20%, meaning that the classifier is very effective in identifying actual true attacks in the dataset. The F1 Score of 93.09% shows a good overall performance of the model. The evaluation metrics for the ANN classifier are shown in Table VIII and the confusion matrix heatmap is in Fig. 7.

- Support Vector Machine:

A well-known classification method that can handle both linear and non-linear datasets is Support Vector Machine SVM. It is founded on the idea of separation between hyperplanes, with SVM's main objective being to find the optimal hyperplane that widens the gap between groups. In general, several kernel functions, ranging from linear to nonlinear kernels, may be utilized to describe the hypersurface. It is an approach to supervised machine learning that may be applied to classification or regression issues. It transforms your data using a method known as the kernel trick and then determines the best boundaries between potential outputs based on these alterations.

TABLE VIII. ARTIFICIAL NEURAL NETWORK PERFORMANCE METRICS

Metric	Value
Accuracy	0.9320
Precision	0.9316
Recall	0.9320
F1 Score	0.9309

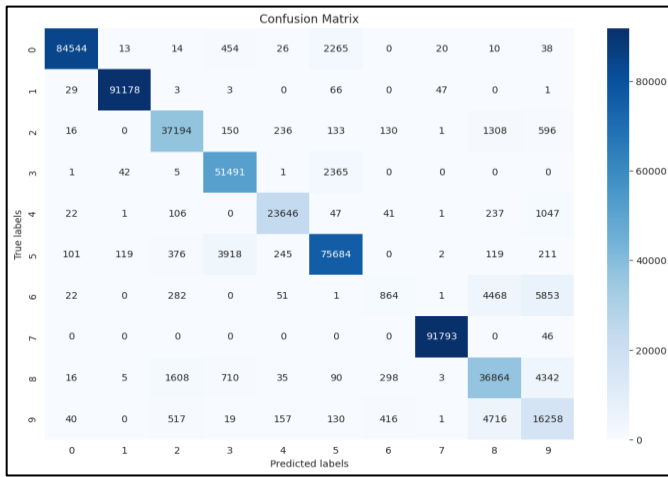


Fig. 7. Confusion matrix heatmap for artificial neural network.

The Support Vector Machine classifier was able to accurately predict 77.09% of the overall instances in the data. In addition, it accurately predicted 79.2% of the True positive instances compared to all of its predicted positive instances (0.79 precision), and it truly identified 77% of the actual positive instances (0.77 recall). Finally, the overall performance was evaluated by the F1 score, achieving a 75.27% value. Table IX shows the performance of the Support Vector Machine classifier and the confusion matrix heatmap for SVM is shown in Fig. 8.

TABLE IX. SUPPORT VECTOR MACHINE (SVM) PERFORMANCE METRICS

Metric	Value
Accuracy	0.7709
Precision	0.792
Recall	0.7709
F1 Score	0.7527

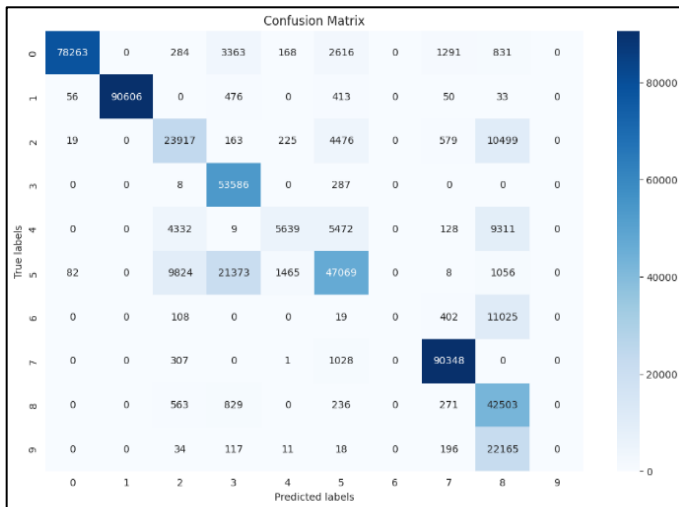


Fig. 8. Confusion matrix heatmap for SVM classifier.

Finally, from the obtained results, the performance of all five classifiers can be compared to selecting the model that is best suited for the task of identifying attacks in the NF-ToN-IoT dataset. Table X shows that the Random Forest RF classifier was able to score the highest accuracy (0.98), followed by the Decision Tree DT classifier (0.97) and the Artificial Neural Network ANN model (0.93). On the other hand, the Support Vector Machine SVM model scored a low accuracy of 0.77, whereas the lowest accuracy was achieved by the Naïve Bayes NB classifier (0.72). As for the other metrics, such as precision and recall, they can be summed up by the F1 score. The highest F1 score was achieved by the Random Forest RF Model (0.98), followed by the Decision Tree DT and Artificial Neural Network ANN (0.97 and 0.93, respectively). A low F1 score was attained by the Support Vector Machine SVM model (0.75), but the lowest F1 score was for the Naïve Bayes NB classifier, where it scored only 0.7051. A visual representation of the performance of the 5 classifiers in terms of accuracy, precision, recall, and F1 score is shown in Fig. 9.

TABLE X. COMPARISON OF THE PERFORMANCE OF THE 5 DIFFERENT CLASSIFIERS ON THE NF-TO-N-IOT DATASET

Classifier	Accuracy	Precision	Recall	F1 Score
<b>NB classifier</b>	<b>0.7275</b>	<b>0.7567</b>	<b>0.7275</b>	<b>0.7051</b>
<b>RF classifier</b>	<b>0.9841</b>	<b>0.9840</b>	<b>0.9841</b>	<b>0.9840</b>
<b>DT classifier</b>	<b>0.9708</b>	<b>0.9706</b>	<b>0.9708</b>	<b>0.9707</b>
<b>ANN classifier</b>	<b>0.9320</b>	<b>0.9316</b>	<b>0.9320</b>	<b>0.9309</b>
<b>SVM classifier</b>	<b>0.7709</b>	<b>0.792</b>	<b>0.7709</b>	<b>0.7527</b>

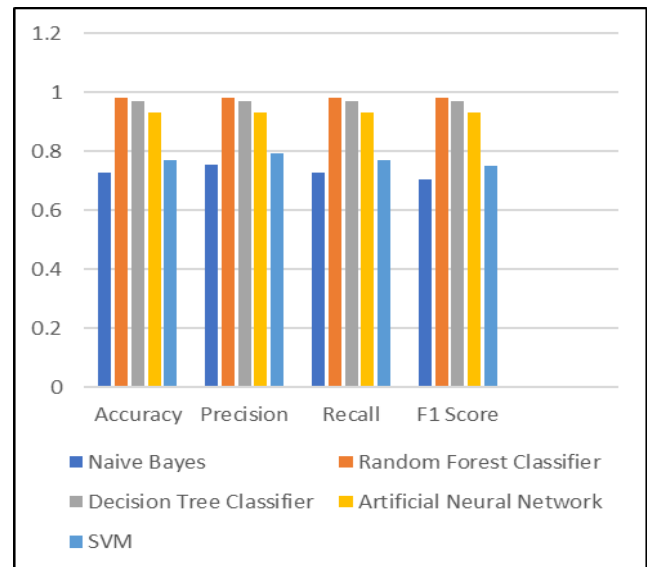


Fig. 9. Comparison of performance of the different classifiers on the NF-ToN-IoT dataset.

## V. RESULT AND DISCUSSION

The inability of standard security systems relying on signatures and rules to identify complex breaches is well recognized. Therefore, utilizing ML and deep learning techniques and various datasets, a number of architectures and algorithms have been created in the prior literature to identify assaults and aberrant behaviors in IoT networks. The studies cited in the literature review were able to achieve relatively good and other not-so-good results. Some studies achieved low accuracy of 85%, 88%, and 89% through RF, DNN, SVM, BFV, and swarm NNs techniques in both [8], [12], and [20]. Other results were able to achieve higher accuracies. For instance, Other studies achieved higher accuracy, ranging from 91% to 92%, by using CNN, RF, and ANN techniques in both [10] and [9], while studies in [11], and [14] achieved an accuracy of 93%. Through IDS and KNN, RF, SVM. On another hand, the studies in both [7] and [13] achieved an accuracy of 94%, which is considered the highest accuracy among the studies through the use of SVM and RF algorithms.

When comparing our methodology, a higher accuracy than all previous studies was achieved by 98% with the RF model.

While it achieved accuracy for both the Decision Tree classifier (0.97) and the artificial neural network model (0.93). On the other hand, the Support Vector Machine model scored a low accuracy of 0.77, while the lowest accuracy was achieved by the Naïve Bayes classified as 0.72. Additionally, the bulk of the researches mentioned in the table used ML and DL systems that were deemed untrustworthy since they were trained mostly on an outdated and unreliable dataset with low accuracy. A more current data set was produced to address this issue and published in [15] [16], which were included in our technique.

The diverse character of the Internet of Things is reflected in this dataset, also known as NF-ToN-IoT. Even though NF-ToN-IoT is better suited for IoT contexts, earlier literature was found to lack data-gathering implementation. Because there aren't many references that utilize the same dataset that we used in our study, a variety of references were employed. A number of references were used that apply to and use different datasets. The following Table XI shows a comparison of our models with other work in the literature review. Also, the visual representation of these results can be seen in Fig. 10.

TABLE XI. COMPARED LITERATURE REVIEW WITH PROPOSED MODEL'S RESULTS

Ref.	Author & year	Study name	Method or Technique	Dataset	Accuracy
[7]	Zhu, Hui et al. <b>2017</b>	Efficient and Privacy-Preserving Online Medical Prediagnosis Framework Using Nonlinear SVM.	ML, SVM	PID	94%
[8]	by Samrat Kumar Dey and Md. Mahbubur Rahman. <b>2019</b>	Effects of Machine Learning Approach in Flow-Based Anomaly Detection on Software-Defined Networking.	ML, RF,DNN	NSL-KDD	82%-88%
[9]	Palla, Tarun Ganesh, and Shahab Tayeb. <b>2021</b>	Intelligent Mirai malware detection for IoT nodes.	ML, ANN, RF	NBaloT	92.8%
[10]	Susilo, Bambang, and Riri Fitri Sari. <b>2020</b>	Intrusion detection in IoT networks using deep learning algorithm.	DL, CNN, ML, RF, MLP	Bot-IoT	88%-91%
[11]	Awajan, Albara. <b>2023</b>	A novel deep learning-based intrusion detection system for IOT networks.	DL, FC, IDS	DID	93.21%
[12]	Ruoli Zhao, Yong Xie, Xingxing Jia,Hongyuan Wang, and Neeraj Kumar. <b>2017</b>	Practical Privacy Preserving-Aided Disease Diagnosis with Multiclass SVM in an Outsourced Environment	SVM, BFV	UCI dermatology	85.4%.
[13]	Alqahtani, Hamed, et al. <b>2020</b>	Cyber intrusion detection using machine learning classification techniques	ML, RF	KDD'99 cup	94%
[14]	Chola, Channabasava, et al. <b>2021</b>	IoT based intelligent computer-aided diagnosis and decision making system for health care.	KNN, SVM, RF, LR and AB	UCI heart disease	93.54%
[20]	Awotunde, Joseph Bamidele, et al. <b>2021</b>	A deep learning-based intrusion detection technique for a secured IoMT system.	IoMT, swarm NNs	NF-ToN-IoT	89%
<b>The proposed Model</b>			<b>RF, NB, DT, ANN, SVM</b>	<b>NF-ToN-IOT</b>	<b>RF 98%</b>

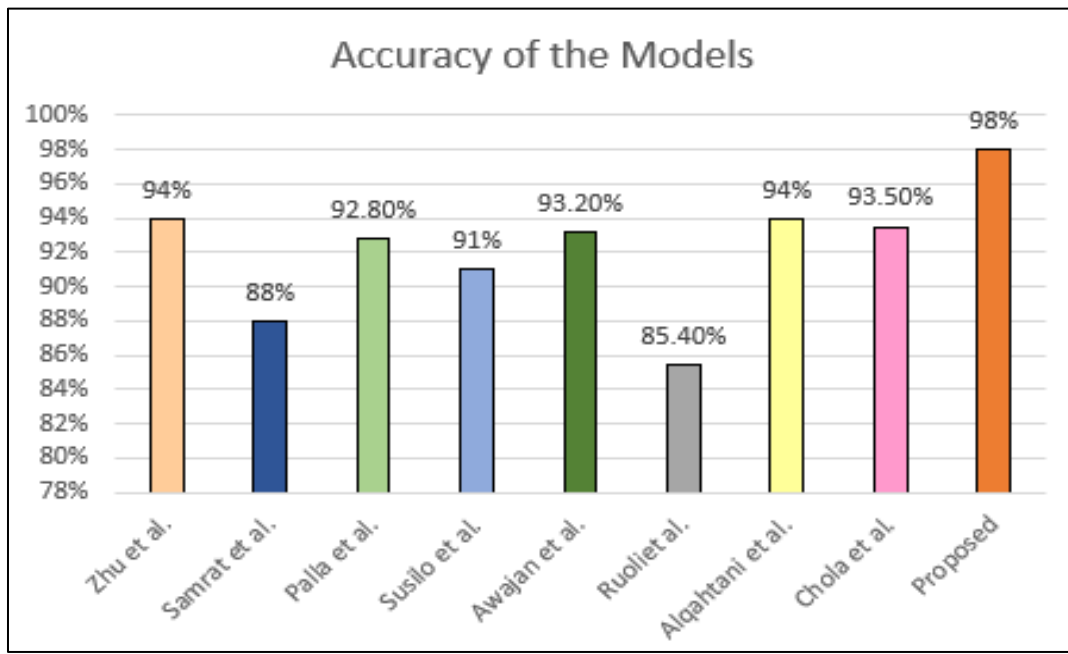


Fig. 10. Histogram showing the Accuracy of Different Models from Literature Review and the Proposed RF Model.

## VI. CONCLUSION

The internet of things IoT environment deals with a huge amount of private and sensitive health data that must be kept safe from tampering or theft. If safety precautions are not implemented, these dangers and assaults against IoT devices in the health sector might completely destroy this industry. These attacks are often carried out to make money, either by selling the stolen data or by holding the victim's data at ransom to release their data. In the healthcare sector, using technology especially IoT technology can be a big leap forward towards providing a better service for patients and facilitating communications and sharing essential files or tasks. For this reason, implementing IoT in healthcare is essential.

However, several privacy and security concerns arise when using IoT. If safety precautions are not implemented, these dangers and assaults against IoT devices in the health sector might completely destroy this industry. These attacks are often carried out to make money, either by selling the stolen data or by holding the victim's data at ransom to release their data.

Thus, it becomes equally important to implement a system that can provide security while implementing IoT. In this research, five different classifiers were employed to predict the occurrence of attacks while using IoT services. For this purpose, the NF-ToN-IoT dataset was selected, where the data were pre-processed before being split into training and testing data. Upon testing the models, it was evident that the RF model achieved the best results; scoring the highest accuracy (0.98) and highest F1 score (0.98). The second-best model was DT classifier, followed by the ANN model.

## VII. FUTURE WORK

The future works aim to experiment with models with different algorithms and different data sets, as well as to

combine several deep and machine learning algorithms, in order to come up with models that give the highest possible accuracy rates and the lowest possible loss rates to obtain the best optimal results in classifying attacks in IoT devices in the electronic health sector and a comparison between data sets and algorithms.

## ACKNOWLEDGMENT

The Vice Presidency for Graduate Studies and Scientific Research at Jouf University is funding this study as a part of its initiative to encourage scientific publications.

## REFERENCES

- [1] Mustafa, Twana, and Asaf Varol. "Review of the internet of things for healthcare monitoring." 2020 8th International Symposium on Digital Forensics and Security (ISDFS). IEEE, 2020.
- [2] Ibrahim, D., and N. Majma. "Improvement of Data Transfer Reliability in IoT-based Coronavirus Patients' Health Monitoring System using by IoT Analytics Expert Systems." CENTRAL ASIAN JOURNAL OF MATHEMATICAL THEORY AND COMPUTER SCIENCES 4.3 (2023): 18-38.
- [3] Kelly, Jaimon T., et al. "The Internet of Things: Impact and implications for health care delivery." Journal of medical Internet research 22.11 (2020): e20135.
- [4] Zeadally, Sherali, et al. "Smart healthcare: Challenges and potential solutions using internet of things (IoT) and big data analytics." PSU research review 4.2 (2020): 149-168.
- [5] Kadhim, Kadhim Takleef, et al. "An overview of patient's health status monitoring system based on internet of things (IoT)." Wireless Personal Communications 114.3 (2020): 2235-2262.
- [6] Alenoghena, Caroline Omoanase, et al. "eHealth: A survey of architectures, developments in mHealth, security concerns and solutions." International Journal of Environmental Research and Public Health 19.20 (2022): 13071.
- [7] Zhu, Hui et al. "Efficient and Privacy-Preserving Online Medical Prediagnosis Framework Using Nonlinear SVM." IEEE journal of biomedical and health informatics vol. 21,3 (2017): 838-850. doi:10.1109/JBHI.2016.2548248

- [8] Dey, Samrat Kumar, and Md Mahbubur Rahman. "Effects of machine learning approach in flow-based anomaly detection on software-defined networking." *Symmetry* 12.1 (2019): 7.
- [9] Palla, Tarun Ganesh, and Shahab Tayeb. "Intelligent Mirai malware detection for IoT nodes." *Electronics* 10.11 (2021): 1241.
- [10] Susilo, Bambang, and Riri Fitri Sari. "Intrusion detection in IoT networks using deep learning algorithm." *Information* 11.5 (2020): 279.
- [11] Awajan, Albara. "A novel deep learning-based intrusion detection system for IOT networks." *Computers* 12.2 (2023): 34.
- [12] Zhao, Ruoli, et al. "Practical Privacy Preserving-Aided Disease Diagnosis with Multiclass SVM in an Outsourced Environment." *Security and Communication Networks* 2022 (2022).
- [13] Alqahtani, Hamed, et al. "Cyber intrusion detection using machine learning classification techniques." *Computing Science, Communication and Security: First International Conference, COMS2 2020, Gujarat, India, March 26–27, 2020, Revised Selected Papers 1*. Springer Singapore, 2020.
- [14] Chola, Channabasava, et al. "IoT based intelligent computer-aided diagnosis and decision making system for health care." *2021 International Conference on Information Technology (ICIT)*. IEEE, 2021.
- [15] T.-I. D. N. Moustafa, 2020, [online] Available: <https://cloudstor.aarnet.edu.au/plus/s/ds5zW91vdgjEj9i>
- [16] D'hooge, S.| L. (2023) NF-ton-IOT, Kaggle. Available at: <https://www.kaggle.com/datasets/dhoogla/nftoniot>
- [17] Sahoo, Kabita, et al. "Exploratory data analysis using Python." *International Journal of Innovative Technology and Exploring Engineering* 8.12 (2019): 4727-4735.
- [18] Khaire, Utkarsh Mahadeo, and R. Dhanalakshmi. "Stability of feature selection algorithm: A review." *Journal of King Saud University-Computer and Information Sciences* 34.4 (2022): 1060-1073.
- [19] Dhal, Pradip, and Chandrashekhar Azad. "A lightweight filter based feature selection approach for multi-label text classification." *Journal of Ambient Intelligence and Humanized Computing* (2022): 1-13.
- [20] Awotunde, Joseph Bamidele, et al. "A deep learning-based intrusion detection technique for a secured IoMT system." *International Conference on Informatics and Intelligent Applications*. Cham: Springer International Publishing, 2021.

# A Multispectral Ariel Image Stitching using Decortification and EEG Signal Extraction Technique

Mukul Manohar S<sup>1</sup>, Dr. K N Muralidhara<sup>2</sup>

Assistant Professor<sup>1</sup>, Professor<sup>2</sup>

Department of Electronics & Communication Engineering-Vemana Institute of Technology  
(Affiliated to VTU, Belagum), Bengaluru, India<sup>1</sup>

Department of Electronics & Communication Engineering-PES College of Engineering, Mandya, India<sup>2</sup>

**Abstract**—UAV Videos and other remote-sensing innovations have increased the demand for multispectral image stitching methods, which can gather data on a broad area by looking at different aspects of the same scene. For large-scale hyperspectral remote-sensing images, state-of-the-art techniques frequently have accumulating errors and high processing costs. However, this research paper aims to produce high-precision multispectral mapping with minimal spatial and spectral distortion. The stitching framework was created in the following manner: First, UAV collects the raw input data, which is then labeled as a signal using a connected component labeling strategy that correlates to each pixel or label using the EEG (Alpha, Beta, Theta, and Delta) technique. Next, the feature extraction process follows a novel decortification Hydrolysis CNN approach which extracts active and passive characteristics. Then after feature extraction, a novel chromatographic classification approach is employed for separating features without overfitting. Finally, a novel yield mapping georeferencing technique is employed for all images stitched together with proper alignment and segmented overlapping fields of view. The suggested deep learning model is an effective method for real-time mosaic image feature extraction which is faster by an average of 11.5 times compared to existing approaches as noted on the samples for experimental analysis.

**Keywords**—EEG signal extraction; feature extraction; image stitching; multispectral image; UAV video

## I. INTRODUCTION

Unmanned Aerial Vehicles (UAVs) fascinated a whole lot of people in many nations as soon as they were developed and have since been widely used in both the military and civilian sectors (such as for mapping, catastrophe monitoring, and crowd monitoring) as well as for in-flight investigation and border patrol. The photographs captured by UAVs have several undesired qualities, like big numbers, short image vicinity, a partly cover degree, and multiple strips, due to the restrictions of flying height and camera focal length. It is crucial to achieve a complete panoramic perspective using quick image fusion techniques to gain more detailed information and broaden the range of vision in many particular tasks [1]. The usage of image fusion technology, which has become a hot research area, is common in multimedia applications including virtual reality, remote sensing image processing, and video surveillance where the watching component has a tiny field of view but the expected observation is large. Users have access to a variety of software applications, like Autopano and Panorama Photo Stitcher, in the market for creating panoramas.

According to the registration strategy, there are four broad groups for picture mosaic techniques [2]. The first technique relies on the intensity levels or colors of each pixel; while this scheme is straightforward, it performs poorly in terms of noise avoidance and blending effectiveness. The second technique relies on the transform features, which has high noise resistivity, but the stitching requires a lot of calculations and the results are subpar when the image changes in view perspective or zoom. The other technique is based on features, for which the stitching efficiency, resilience, and accuracy of the algorithms are often high and they are typically similar to image size editing, transformation, and rotation. The three conventional feature extraction algorithms are Harris, Scale-Invariant Feature Transform [3], and Oriented Fast and Rotated Brief. On deep learning, the final one is based. Many deep learning methods based on feature extraction and matching have been created to perform picture registration [4]. Despite the recent advancements in image mosaic technology, several techniques still fall short of the instantaneous, reliable, and accuracy demands of UAV image fusions. UAV pictures have a lot of data, tiny phase amplitude, and a higher degree of overlapping [5], among other qualities. Since a UAV video/image contains a lot of data, the mosaic process takes an unacceptably long time [6]. Second, because they are frequently small, UAVs struggle to maintain themselves and are poorly wind-resistant. There will unavoidably be some tilting when taking pictures, even though they have an autopilot and a stabilizing gyroscope. Images captured by UAVs contain significant affine distortion in comparison to the actual scene due to the geometric distortion of the camera caused by the lens [7]. To address this problem, researchers have proposed mounting high-resolution cameras on a UAV. The dimensional fault, which is the difference between the relative offset and the absolute offset, can be thought of as low-frequency noise brought on by the UAV drift, and it can be addressed by using a high-pass filter [8].

The projection matrix-based approach can also be used to extract supreme structural offset from the drone footage, presuming that out-of-plane offsets can be disregarded. The projection matrix was determined using stationary backdrop features, and then it was further honed using the constrained bundle adjustment optimization approach to reduce the re-projection error. Additionally, for this technique to work, the camera must move rather quickly; otherwise, the camera settings in successive frames would be very much close, which will cause a major inaccuracy when applying the bundle

adjustment. However, in-field measurements, the expanse between immobile locale objects and the drone is typically many times greater than the drone motion itself, thus making this approach inapplicable. With the help of homography, also referred to as perspective transformation, it is possible to adjust the camera movement for a planar outlook without the use of camera constraints [9].

Although the homography related methods are numerically effective, there are numerous obstacles in the way of its straight-forward application to the counting of dynamic offsets of massive structures. Because of the limited pixel resolution, it's possible that the camera won't be able to sufficiently capture both the offset measurement points and immobile areas of the organization (i.e., the homography characteristics) in a single image. However, choosing at random from these areas could cause big inaccuracies in the homography computation result. The chosen homography characteristics is expected to be over or near the flat surface in which structural motion occurs since homography requires a planar scene and not all random homographies are realistic picture alterations [10]. And the system's created aerial panorama could have poor aesthetic flaws at pivot points. Then, the stitching is successful for neighboring frames in image sequences where the stitching has been successful in order, but it can be challenging to guarantee the overall stitching effect in multistrip long-distance trips [11]. Additionally, it is challenging to produce useful data from sequences gathered by UAVs during multistrip missions and it is laborious to manually track the chosen homography elements in each frame of a video. The stitching effect, however, can be diminished with a greater volume of images and much identical parts in the prospect [12]. However, it is difficult finding a stationary landmark during an earthquake.

In this work proposed, a novel live drone image mosaic system is proposed. The main contributions are as follows:

- First, proposed a state-of-the-art, real-time framework for drone image mosaicking. The architecture comprises automatic initialization, feature extraction & classification, and live mosaic generation. The setup procedure uses EEG to automatically identify the image clips based on brainwaves (Alpha, Beta, Theta, and Delta).
- After that, all four of these brainwaves are subjected to a feature extraction procedure utilizing deep learning, where a mixture of alpha and beta is used to examine active features while theta and delta are used to consider passive features.
- After feature extraction, the classification process is done then finally real-time mosaic creation makes up the framework's entire structure.
- The primary tasks required in the controlling of large areas, including mosaicking, feature extraction, and classification, are improved in accuracy compared to existing drone-based techniques.

The remaining portion of the paper is structured as follows: Section II explains briefly the existing literature and the research gaps. Section III gives an overview of the proposed

work. Section IV details the proposed system architecture. Section V discusses the results and analysis of implementation. Section VI finally concludes the work.

## II. RELATED STUDY

There has been interesting works available in the literature that paves way for the growth of the techniques and the technology responsible for image applications. Avola et al. [13] proposed the best parameters to automatically predict, specifically, depth and frame rate to recognize the three factors in previous section are suggested and tested. The dimensions of the aimed object to be analyzed, the UAVs' travel speed, and the primary inner dimensions of the video sensor, such as the focal length, field of view, and pixel size, are some of the criteria that are used to estimate the parameters. Both man-made videos produced with the in-flight data and Robotics Simulation (AirSim) and actual film sequences reported in the UAV Mosaicking and Change Detection (UMCD) and NPU Drone-Map datasets served to demonstrate the suggested method's complete effectiveness on the objective. However, ensuring the minimal spatial resolution necessary to complete a given activity is the primary issue to be addressed regarding flying height.

Ranghao et al. [14] proposed a live drone image fusion architecture that solely utilizes the UAV picture frames and does not depend on the global positioning system (GPS), ground control points (CGPs), or any other auxiliary data. Through the use of this framework, it is hoped to produce high-quality panoramas while reducing spatial distortion and speeding up mosaicking operations before choosing key frames to increase efficiency, the framework evaluates the general setting of every new frame to be added. Then, to perform an accurate position computation of the existing scene and lessen the bend brought on by increasing mistakes, a new optimization method based on minimizing weighted reprojection errors is implemented. To produce the best mosaic output, the local picture is fused and updated in real-time using the weighted partition fusion approach based on the Laplacian pyramid. UAVs frequently capture multi-strip and large-scale image sequences, however, it is challenging to create panoramas directly from the stitched photos, and some of the feature points are unstable and challenging to extract.

Srivastva et al. [15] proposed an overview of deep learning methods for on-ground vehicle recognition utilizing aerial data obtained by UAVs (often referred to as drones). To review the works, it is important to consider both the optimization goal and the method used to increase accuracy and decrease computation overhead. To illustrate the parallels and discrepancies between different approaches and to draw attention to the remaining issues in this field, this work is a useful study. Researchers studying AI, traffic surveillance, and UAV applications will find this survey to be useful. However, the processing of a picture takes a long time and results in a lot of false positives.

Woo et al. [16] proposed to drastically lower the error level, identification of fissures was determined using comparative position between components in drone-captured photos rather than using absolute position information. A total of 97 photos were collected using aerial photography. Five



fissures and three reference components were defined using the point-cloud approach, image blending, and homography domain algorithm. Importantly, the comparison of calculated localized values with concrete values obtained from field measurement showed that errors in the range of 24-84 mm and 8-48 mm, respectively, were received based on the coordinates. Additionally, RMSE errors ranging from 37.95 to 91.24 mm were verified. The target concrete construction, however, might not always have enough or the right reference objects accessible.

Rui et al. [17] proposed a Rapid Scale-Invariant Feature Transform (RSIFT) operator to shorten computation time. Then used is the As-Natural-As-Possible (AANAP) technique for picture registration. To remove the motion ghosting, an image segmentation method is used. Finally, a unique collection of aerial photos is created for image mosaics, using which analogous tests using cutting-edge image mosaic techniques are carried out. However, research indicates that SIFT has huge computational complexity.

Zhang et al. [18] proposed an approach based on Oriented Fast and Rotated Brief (ORB) and semantic segmentation. To distinguish between the forefront and surroundings of the image and to get the forefront content, a semantic segmentation network is introduced by the algorithm during image registration. It simultaneously extracts feature points using the quadtree decomposition concept and the conventional ORB approach. The foreground feature points can be removed to achieve feature point matching by comparing the feature point information with the foreground semantic information. The homography domain and the weighted fusion technique will be used to stitch and fuse the images based on precise image registration. However, it is challenging to find trait points in areas with weedy texture, and the major drawback of SIFT is that real-time performance cannot be attained without the aid of hardware or specialized image processors.

In order to locate an intuitively optimal seam in uninterrupted areas of high similarity, Yaun et al. [19] suggested an innovative super pixel-based performance function that incorporates data on material difficulty, color distinction, and gradient difference. Finally, to eliminate seams and provide seamless color transitions, employ an exceptional pixel-based color blending technique. The method is superior to several cutting-edge UAV photo stitching techniques, according to experimental data, and can effectively and rapidly perform seamless stitching. However, owing to the existence of parallax, artifacts always show up in overlapped areas.

Yang et al [20] proposed a straightforward yet effective approach for sewing durable and precise blade images. Introduce the Blade30 dataset, which includes 1,302 real drone-captured photos of 30 entire blades taken in a variety of environments (both on and off-shore) and richly annotated with information on flaws and contaminations, among other things, to encourage further research. By using the recommended patching strategy based on drone-blade lengths and rotor margins at the coarse-grained level, the initial blade portrait is produced. Then, utilizing regression-based roughness and shape losses, fine-grained modifications are optimized. Additionally, this approach makes full use of the drone's

existing knowledge and the characteristics of blade pictures. However, it may be challenging to gather enough labeled data to initiate this, which is primarily due to inefficient inspection techniques.

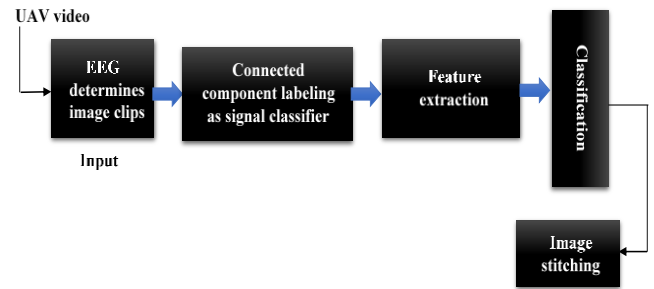


Fig. 1. Block diagram for real-time image stitching of UAV video.

Some of the research gaps identified are: Low-textured picture sewing still presents a number of complicated challenges; since multiple photos were captured from various angles, the patched images also contain projective distortions; Picture stitching combines a number of overlapping images to create a single, larger mosaic with a wider angle of perspective because cameras have a low angle of perspective. Global geometric transformations were typically approximated by older methods for stitching together overlapping pictures; These methods, however, rely on rigid premises that are commonly violated in reality, resulting in defects in the merged images like misalignments or ghosting, such as the camera rotation having a fixed projective center or the scene having a constrained depth variance; Only a small number of points are recognized and matched in some homogeneous regions, making it challenging to predict an exact transformation. The proposed method tries to overcome these challenges and develops an efficient method for image stitching.

### III. OVERVIEW OF THE PROPOSED WORK

The suggested architecture seeks to address the several issues mentioned above for improved target surveillance and real-time image stitching of UAV video. Enhancing real-time target tracks and stitching image serves several purposes, including improved accuracy, easier long-distance stitching, accurate information, and overall performance. It also captures both displacement measurements and homography properties in a single image. UAVs, such as drones or planes, equipped with cameras, sensors, software for control, and interactions, first gather raw data. UAVs can collect visual sensing data via their camera-equipped device. [21]. The pixels or labels in these images are designated as signals, utilizing an analogy to EEG patterns such as alpha, beta, delta, and theta, each linked to distinct frequencies [22]. The process of extracting active and passive features utilizing statistical methodology to ascertain the average weight for all waves and set thresholds is then carried out. If the weighted average value is higher than the threshold, it is considered active else passive. Algorithms based on CNN relate active attributes to alpha and beta, whereas passive attributes are linked to delta and theta. However, this method permits accurate feature extraction,

captures the qualities of both displacement measurement and homography, and has superior accuracy. After this, the classification process follows, effectively segregating these features to prevent overfitting. Finally, using overlapping fields of view, the images are seamlessly combined to provide a segmented, high-resolution image. However, this method makes stitching across large distances simple, provides accurate information, and enhances overall performance. Fig. 1 depicts a block diagram for real-time image stitching of UAV video, which demonstrates raw data. Video is acquired by UAV like a drone or plane, and comprises of a sequence of image clips assessed using EEG. Images are categorized as signals using the connected component labeling technique. Then feature extraction and classification are performed. Finally, the images are seamlessly stitched together to produce a cohesive output.

#### IV. ARCHITECTURE OF THE PROPOSED METHODOLOGY

The proposed design solves low accuracy, overfitting difficulty ensuring stitching effect over long distances, incorrect data, and difficulty capturing both displacement and homography characteristics in a single image. In the beginning, raw data is collected by UAV to capture illustrated sensing information by its camera installed, and the pixels or labels in these images are designated as signals using the connected component labeling technique for using an analogy to EEG patterns such as alpha, beta, delta, and theta, where each linked to distinct frequencies. For removing noise or error in the signal and the signal is then split into two halves using decortification CNN, a proposed feature extraction method that collects both active and passive features. While delta and theta represent passive features, alpha, and beta indicate active features. To identify the alpha, beta, delta, and theta features, use a statistical method. To ensure this, average weights for each wave must be determined, and global thresholds must be set. The characteristic is regarded as active if the estimated weighted average value is higher than the threshold; otherwise, it is regarded as passive.

After feature extraction, classification is carried out by employing the chromatographic approach to separate active and passive features separately without overfitting. However, this approach enables precise feature extraction and categorization and captures both the characteristics of displacement measurement and homography. Finally, a high-resolution image is created by combining all the photographs with their overlapping fields of view, utilizing a yield mapping geo-referenced approach to appropriately align and segment each image. However, this approach makes stitching across vast distances straightforward, high accuracy offers correct data, especially finding stable landmarks after an earthquake is simple and improves overall performance.

The proposed architectural model for the real-time UAV video image stitching process is shown in Fig. 2 which start with the UAV obtaining the raw data and this data comprise a sequence of image clips. Each pixel/label within these images is categorized as a 'signal' using a connected component labeling technique. This idea makes use of the four types of brain waves alpha, beta, theta, and delta to extract active and passive features using CNN, based on their respective

frequencies. The active attributes encompass alpha and beta waves, while the passive attributes encompass theta and delta waves. Subsequently, a classification process utilizing chromatography technique is applied to the extracted features. Finally, the process involves image stitching using a yield mapping geo-referenced technique, seamlessly combining the acquired images.

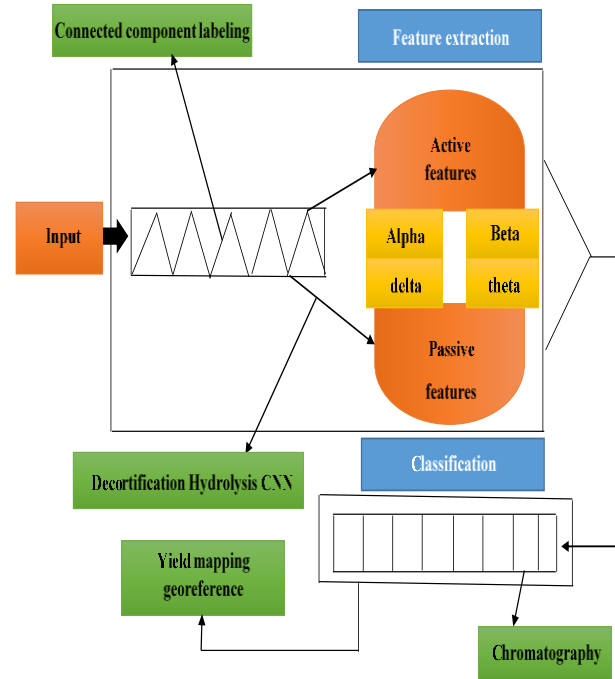


Fig. 2. Proposed architectural model for Real-Time UAV video image stitching.

All of these EEG signal properties (Alpha, Beta, Theta, and Delta) are described below, along with their mathematical expressions:

Average frequency weighting (AFW): The average frequency weighting can be found to estimate the power spectrum of the obtained sequence of signal [22]. Using statistical methods an average value is found for all waves with certain frequency weighting elements. The following mathematical equation can be used to represent the AFW,  $S(k)$  for any N-point signal  $x(n)$ .

$$S(k) = \frac{1}{N} * W \sum_{K=0}^{N-1} |X_p(k)| \quad (1)$$

where  $n$  is a variable that specifies the data points of the time-domain discrete-time input signal  $x(n)$ . For a total number of temporal data points  $N, 0 \leq n \leq N-1$ . Similar to that,  $k$  also refers to a variable that specifies data points in spectral-domain form  $X_p(k)$  of  $x(n)$ . For a total number of spectral data points  $N, 0 \leq k \leq N-1$  [23].

Repeating this calculation across all frequency bands yields the final Power Spectral Density (PSD) values for Delta, Theta, Alpha, and Beta [24].

The PSD according to Welch is expressed by:

$$P_d(f) = \frac{1}{MU} \left| \sum_{k=0}^{N-1} X_p(k)w(k)e^{-j2\pi fk} \right|^2 \quad (2)$$

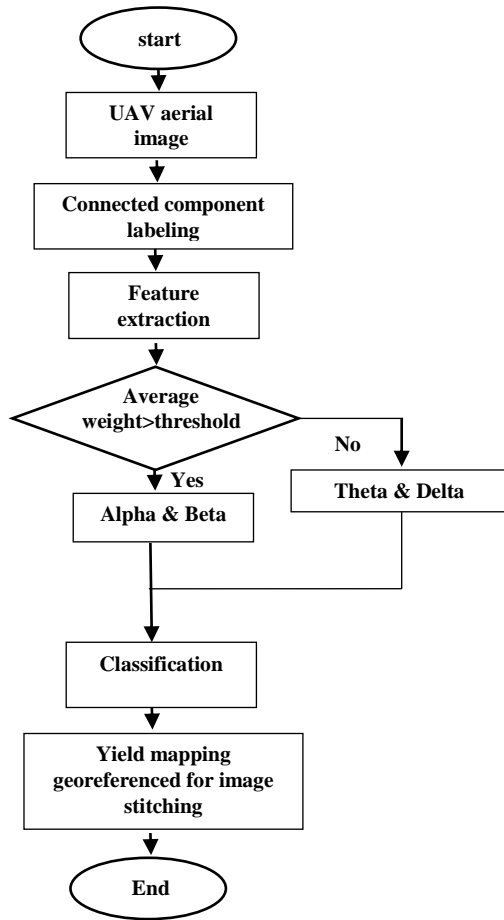


Fig. 3. Flow chart of the proposed architecture.

Let  $x(n)$  be the sequence, with  $N = 1, 2, 3, \dots, N-1$  representing the signal intervals, and  $M$  denoting each interval length.  $U$  is the normalization factor for the power in the window function, and  $w(n)$  is the windowed data, so:

$$U = \frac{1}{M} \sum_{k=0}^{N-1} |w(n)|^2 \quad (3)$$

After extracting all PSD, the alpha, and beta signals are considered active and delta and theta are considered passive features to compare with the overall threshold  $\emptyset$

$$\text{PSD} > \emptyset \text{ or } \text{PSD} < \emptyset \quad (4)$$

The suggested architecture's flowchart is shown in Fig. 3, starting with raw data which is an aerial image that is collected by an UAV as input. A connected component labeling technique is used to classify each pixel or label in these images as a "signal". This concept uses the four different types of brain waves alpha, beta, theta, and delta to extract both active and passive properties. Using a statistical approach, the alpha, beta, delta, and theta features are identified and designated as  $B_1, B_2, B_3, B_4$ , respectively. To do this, average wave weights and global thresholds must be established. The attribute is considered active if the projected weighted average value

exceeds the threshold; otherwise, it is considered passive. The procedure of classification is then performed following feature extraction. Finally segmented using yield mapping georeferenced approach and merged every image with proper alignment.

**Proposed Algorithm:**

Multispectral mapping for image stitching using EEG signal to extract robust feature extraction

**Require:** Multispectral images  $I_p = 1, 2, 3 \dots N$

**Ensure:** Multispectral mapping with high alignment accuracy

**Step 1:** Start  $I_p = 1, 2, 3 \dots N$

**Step 2:** Each pixel in the image is labeled as a signal.

**Step 3:**  $i=1$

**Step 4:** for each sample,  $S_m$  do

**Step 5:** Calculate the average weighted probability for each wave  $B_1, B_2, B_3, B_4$ , and set the threshold  $\emptyset$ .

**Step 6:** Compare weighted probability with threshold

**Step 7:** if  $B > \emptyset$  or  $B < \emptyset$  then

```

{
incorporate  $B_1, B_2$ ;
else {
incorporate  $B_3, B_4$ ;
}
}
i++
end for
    
```

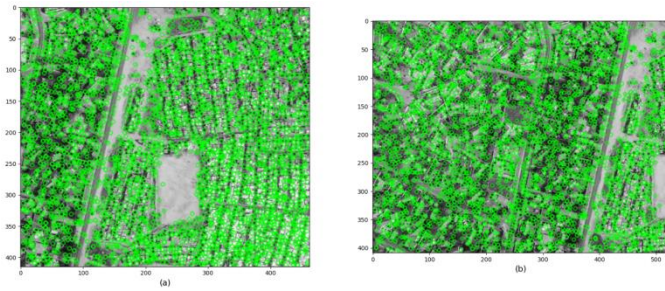
V. RESULT AND DISCUSSION

This section presents the detailed results of the proposed model. To provide a comprehensive performance evaluation, the suggested model was compared to two other models, the SIFT [25] and BRISK [26]. The work was implemented in Python 3.11.4 and the packages used are cv2 and math. Two sample images were used to test the performance of the proposed work.

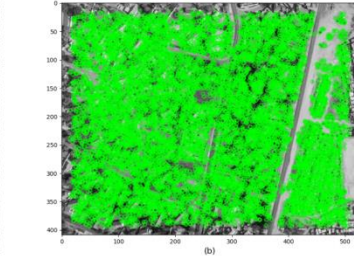
Fig. 4 shows the input sample 1 image pairs to be stitched. During feature extraction using SIFT algorithm, 4369 features from the first image and 4333 features from the second image are selected as shown in Fig. 5(a) using SIFT. Fig. 5(b) shows the 10038 features and 10774 features selected from images 1 and 2 respectively using BRISK algorithm. Fig 5(c) shows the 500 features from Image 1 and 500 features from image 2 selected using the proposed approach. Fig 6(a), (b) and (c) shows the matching features between the pairs of images using SIFT, BRISK and the proposed algorithm respectively. Fig. 7 shows the sample 1 image pairs fused.



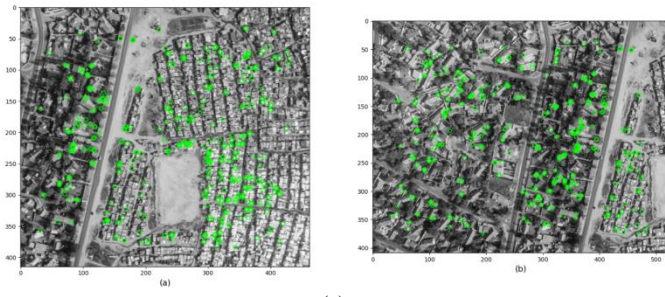
Fig. 4. Input Sample 1: Image1 and Image2.



(a)



(b)



(c)

Fig. 5. Features Extracted (a) SIFT (b) BRISK (c) Proposed.



(c)

Fig. 6. Matching features (a) SIFT (b) BRISK (c) Proposed.



Fig. 7. Final stitched image.

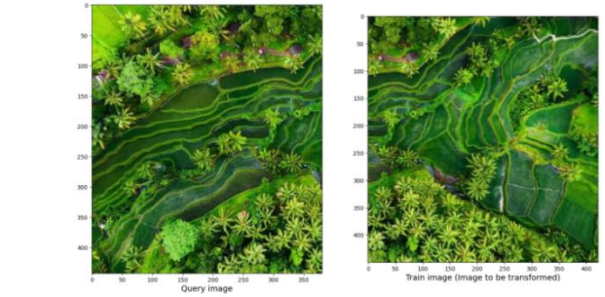
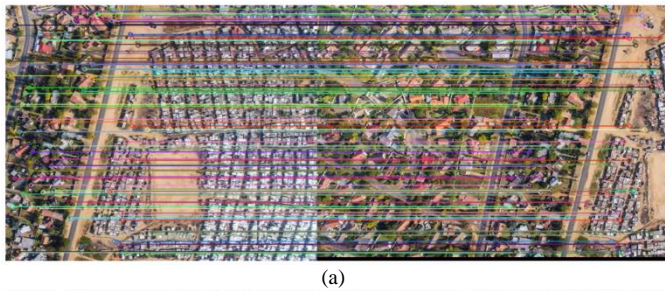
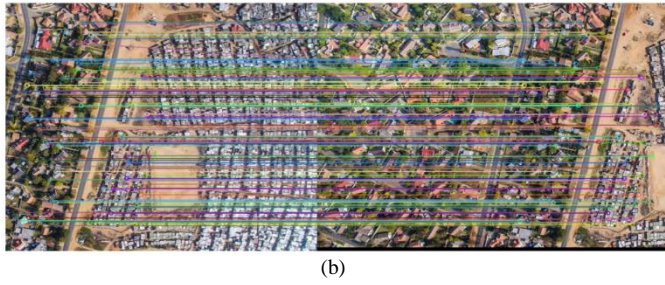


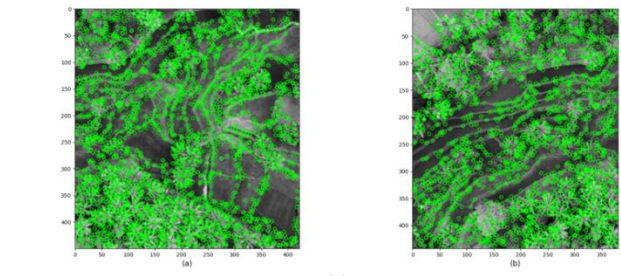
Fig. 8. Input sample 2: image1 and image2.



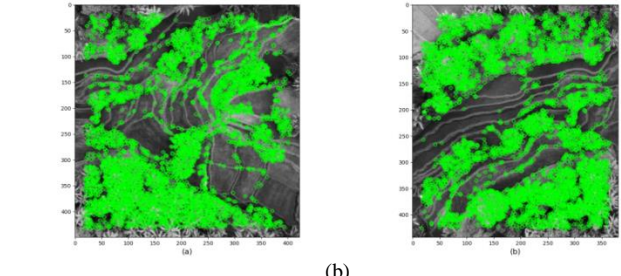
(a)



(b)



(a)



(b)

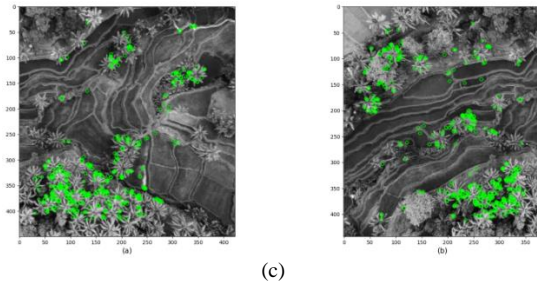


Fig. 9. Features extracted (a) SIFT (b) BRISK (c) Proposed.

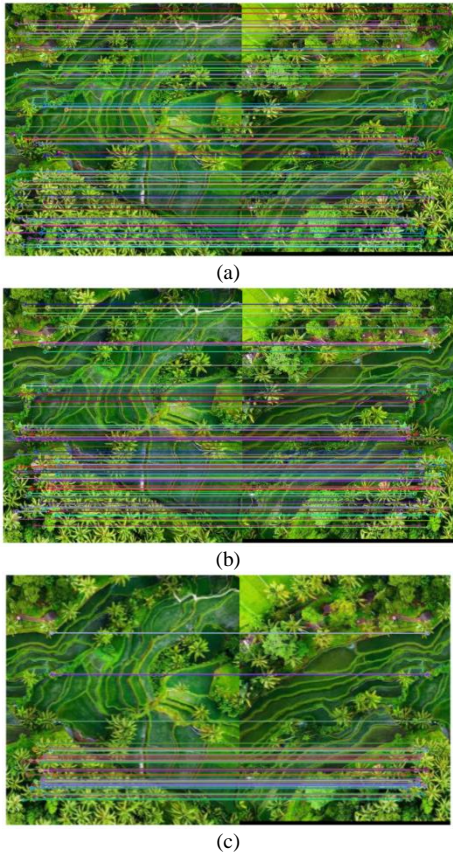


Fig. 10. Matching features (a) SIFT (b) BRISK (c) Proposed.

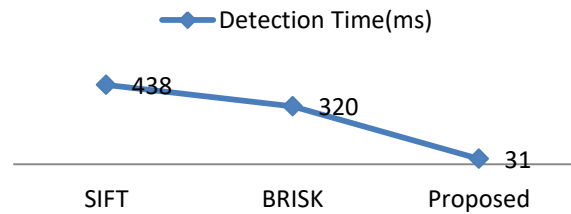


Fig. 11. Final stitched image.

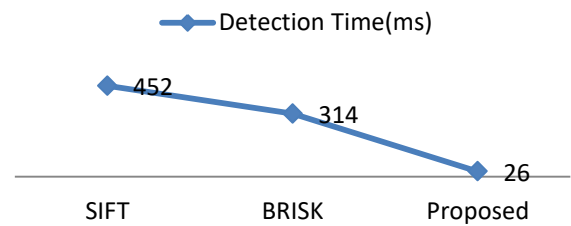
Fig. 8 shows the input sample 2 image pairs to be stitched. During feature extraction using SIFT algorithm, 3532 features from the first image and 2754 features from the second image are selected as shown in Fig. 9(a) using SIFT. Fig. 9(b) shows the 5057 features and 5128 features selected from images 1 and 2, respectively using BRISK algorithm. Fig. 9(c) shows the

475 features from Image 1 and 480 features from image 2 selected using the proposed approach. Fig. 10(a), (b) and (c) shows the matching features between the pairs of images using SIFT, BRISK and the proposed algorithm respectively. Fig. 11 is the stitched image of sample 2 image pairs.

Fig. 12(a) and (b) shows the detection time for feature selection for image mapping for sample 1 image pairs. Fig. 13(a) and (b) shows the detection time for feature selection for image mapping for sample 2 image pairs. The detection time for feature extraction clearly represents that the proposed work is 14 times and ten times faster than the SIFT and BRISK algorithms for the image 1 of sample 1; 17 times and 12 times faster than the SIFT and BRISK algorithms for the image 2 of sample 1; 11 times and nine times faster than the SIFT and BRISK algorithms for the image 1 of sample 2; 9 times and ten times faster than the SIFT and BRISK algorithms for the image 2 of sample 2. Overall, the proposed algorithm on average is 13 times faster than the SIFT algorithm and ten times faster than the BRISK algorithm.

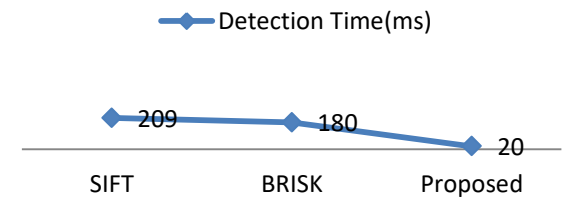


(a)

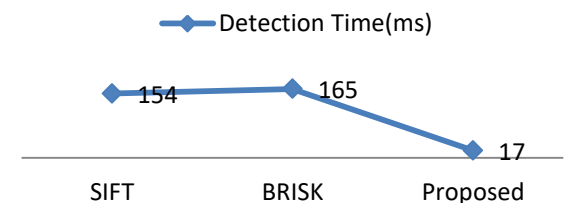


(b)

Fig. 12. Detection times of sample 1 image pairs.



(a)



(b)

Fig. 13. Detection times of sample 2 image pairs.

## VI. CONCLUSION

In this work we have proposed a new technique for image fusion using signal extraction technique behind the brain waves using EEG. Decortification technique is used in processing and preparing the images for matching and blending. Live UAV video frames are fused and can be used for various applications like target tracking and distance measurement. The proposed work is able to choose optimal features with good image understanding. The suggested approach outperforms the SIFT and BRISK algorithms on average by a factor of 13 and 10, respectively. In future the proposed work will be implemented in test-bed for live frame sequence transformations to fused images.

## REFERENCES

- [1] Mhangara, P., Mapurisa, W., & Mudau, N., "Comparison of Image Fusion Techniques Using Satellite Pour P", *Observation de la Terre (SPOT), 6 Satellite Imagery, Applied Sciences*, 2020.
- [2] Jung, H., Kim, Y., Jang, H., Ha, N., & Sohn, K., "Unsupervised Deep Image Fusion With Structure Tensor Representations", *IEEE Transactions on Image Processing*, 29, 3845-3858, 2020.
- [3] Dogra, A., Goyal, B., & Agrawal, S., "Multi-Scale Decomposition to Non-Multi-Scale Decomposition Methods: A Comprehensive Survey of Image Fusion Techniques and Its Applications", *IEEE Access*, 5, 16040-16067, 2017.
- [4] Haskins, G., Kruecker, J., Kruger, U., Xu, S., Pinto, P., Wood, B., & Yan, P., "Learning deep similarity metric for 3D MR-TRUS image registration", *International Journal of Computer Assisted Radiology and Surgery*, 14, 417-425, 2018.
- [5] Yurduseven, O., Fromenteze, T., & Smith, D., "Relaxation of Alignment Errors and Phase Calibration in Computational Frequency-Diverse Imaging using Phase Retrieval", *IEEE Access*, 6, 14884-14894, 2018.
- [6] He, J., Sun, M., Chen, Q., & Zhang, Z., "An improved approach for generating globally consistent seamline networks for aerial image mosaicking", *International Journal of Remote Sensing*, 40, 859 – 882, 2018.
- [7] Zhaoxiang, Z., Iwasaki, A., & Xu, G., "Attitude Jitter Compensation for Remote Sensing Images Using Convolutional Neural Network", *IEEE Geoscience and Remote Sensing Letters*, 16, 1358-1362, 2019.
- [8] Su, Z., Wang, F., Xiao, H., Yu, H., & Dong, S., "A fault diagnosis model based on singular value manifold features, optimized SVMs and multi-sensor information fusion", *Measurement Science and Technology*, 2020.
- [9] Guan, B., Zhao, J., Li, Z., Sun, F., & Fraundorfer, F., "Relative Pose Estimation With a Single Affine Correspondence", *IEEE Transactions on Cybernetics*, 52, 10111-10122, 2021.
- [10] DeTone, Daniel, Tomasz Malisiewicz, and Andrew Rabinovich. "Superpoint: Self-supervised interest point detection and description." *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2018.
- [11] Moniruzzaman, M. D., et al. "Teleoperation methods and enhancement techniques for mobile robots: A comprehensive survey." *Robotics and Autonomous Systems*, 150, 103973, 2022.
- [12] Nie, Lang, et al. "A view-free image stitching network based on global homography." *Journal of Visual Communication and Image Representation* 73,102950, 2020.
- [13] Avola, Danilo, et al. "Automatic estimation of optimal UAV flight parameters for real-time wide areas monitoring." *Multimedia Tools and Applications* 80 (2021): 25009-25031.
- [14] Li, Ronghao, et al. "A Real-Time Incremental Video Mosaic Framework for UAV Remote Sensing." *Remote Sensing* 15.8, 2127, 2023.
- [15] Srivastava, Srishiti, Sarthak Narayan, and Sparsh Mittal. "A survey of deep learning techniques for vehicle detection from UAV images." *Journal of Systems Architecture* 117 (2021): 102152.
- [16] Woo, Hyun-Jung, et al. "Localization of Cracks in Concrete Structures Using an Unmanned Aerial Vehicle", *Sensors* 22.17, 6711, 2022.
- [17] Rui, Ting, et al. "Research on fast natural aerial image mosaic", *Computers & Electrical Engineering*, 90107007, 2021.
- [18] Zhang, Gengxin, et al. "UAV low-altitude aerial image stitching based on semantic segmentation and ORB algorithm for urban traffic." *Remote Sensing* 14.23, 6013, 2022.
- [19] Yuan, Yiting, Faming Fang, and Guixu Zhang. "Superpixel-based seamless image stitching for UAV images." *IEEE transactions on geoscience and remote sensing* 59.2 (2020): 1565-1576, 2020.
- [20] Yang, Cong, et al. "Towards accurate image stitching for drone-based wind turbine blade inspection." *Renewable Energy* 203 (2023): 267-279.
- [21] Do, Hai T., et al. "Energy-efficient unmanned aerial vehicle (UAV) surveillance utilizing artificial intelligence (AI)." *Wireless Communications and Mobile Computing*, pp 1-11, 2021.
- [22] Subha, D.P., Joseph, P.K., Acharya U, R. et al., *EEG Signal Analysis: A Survey*. *J Med Syst* 34, 195–212, 2010.
- [23] Singh, Kuldeep, Sukhjeet Singh, and Jyoteesh Malhotra. "Spectral features based convolutional neural network for accurate and prompt identification of schizophrenic patients." *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine* 235.2, 167-184, 2021.
- [24] Wen, Tee Yi, and SA Mohd Aris. "Electroencephalogram (EEG) stress analysis on alpha/beta ratio and theta/beta ratio." *Indones. J. Electr. Eng. Comput. Sci* 17.1, 175-182, 2020.
- [25] A. Annis Fathima, R. Karthik, V. Vaidehi, *Image Stitching with Combined Moment Invariants and Sift Features*, *Procedia Computer Science*, Volume 19, Pages 420-427, ISSN 1877-0509, 2013.
- [26] Yanli Liu, Heng Zhang, Hanlei Guo and Neal N. Xiong. "A FAST-BRISK Feature Detector with Depth Information", *Sensors (Basel)*, Nov 18(11): 3908, 2018.

# Texton Tri-alley Separable Feature Merging (TTSFM) Capsule Network for Brain Tumor Detection

Vivian Akoto-Adjepong, Obed Appiah, Peter Appiahene, Patrick Kwabena Mensah

Department of Computer Science and Informatics-University of Energy and Natural Resources, Sunyani, Ghana

**Abstract**—Brain tumors represent one of the most perilous and lethal forms of tumors in both children and adults. Early detection and treatment of such malignant disease types may reduce the mortality rate. However, manual procedures can be used to diagnose such disorders, and this process necessitates a careful, in-depth analysis which is prone to errors, tedious for health professionals, and time-consuming. Therefore, this research aims to design a Texton Tri-alley Separable Feature Merging (TTSFM) Capsule Network based on dynamic routing, suitable for the automatic detection of brain tumors. The TTSFM Capsule Network's Texton layer helps to extract important features from the input image, and the separable convolutions coupled with the use of fewer filters and kernel sizes help to reduce the time for training, the size of the model on disk, and the number of trainable parameters generated by the model. The model's evaluation results on the brain tumor dataset consisting of four classes show better performance than the traditional capsule network, and are comparable to the state-of-the-art models, with an overall accuracy of 97.64%, specificity of 99.24%, precision of 97.43%, sensitivity of 97.45%, f1-score of 97.44%, ROC rate of 99.50%, PR rate of 99.00%. The components and properties of the proposed model make the model deployable on devices with low memory like mobile devices. This model with better performance can assist physicians in the diagnosis of brain tumors.

**Keywords**—Texton; separable convolutions; capsule neural network; dynamic routing; brain tumor; brain tumor detection

## I. INTRODUCTION

Brain tumor is among the most fatal and dangerous tumors in both children and adults [1]. Brain and spinal cord tumors are assemblages of abnormal cells that have multiplied uncontrollably inside the brain or spinal cord. There will be 25,050 diagnoses of malignant brain and spinal tumors by 2023 in both men and women[2].

Medical imaging plays a vital role in the diagnosis, monitoring of tumor progression, and treatment of tumors. Magnetic Resonance Imaging (MRI) is the preferred technique for imaging due to its non-ionizing nature. It offers significant insights into the characteristics, dimensions, form, and positioning of brain tumors.

Manually evaluating MRIs is laborious and error-prone, hence an Artificial intelligence (AI)-driven system that operates automatically is required to aid in medical diagnosis. Techniques rooted in machine learning, like support vector machines, have been utilized to aid in accurately detecting

medical conditions [3]. Nevertheless, the outcomes of these approaches fell short of established benchmarks, and the process of extracting features is notably time-intensive. To tackle these challenges, deep learning techniques like convolutional neural networks (CNNs) were embraced to enhance the process of extracting features. Remarkably, CNNs demonstrated a level of performance that is comparable to that of human experts.

Despite CNN's strong achievements, the study found specific constraints including the need for extensive datasets, high computational demands [4], translational invariance [5], and adherence to particular criteria for optimal feature selection [6]. In the field of health, obtaining a voluminous dataset poses a significant hurdle, compounded by a scarcity of skilled annotators and privacy issues [4]. Consequently, to mitigate the overfitting of CNNs on these limited datasets, methods of data augmentation are employed. However, it should be noted that these data augmentation techniques are both time-consuming and labor-intensive [7].

Capsule Network (CapsNet) was introduced to tackle the issues of CNN [8]. In contrast to CNNs, CapsNet does not necessitate extensive datasets, and is resistant to uneven class distributions and spatial orientation changes. These properties of CapsNet render it appropriate for medical image diagnosis. However, CapsNets do have their own set of limitations[9]. They exhibit suboptimal performance on complex images, and those with diverse backgrounds, and try to account for every element in an image. As a result of these properties, the performance of the network may suffer when dealing with detailed malignant images.

In order to further improve CapsNets textural, color, and spatial recognition capabilities, this paper adopts CapsNets dynamic routing algorithm and implements a Texton layer [10], separable convolutions, and a max-pooling layer. This allows CapsNet to decide on which features are essential and the coupling coefficients that need to be decreased in enhancing the hierarchical relationship of closely related capsules. The Texton Tri-alley Separable Feature Merging (TTSFM) Capsule Network model proposed helps to address the crowding problem in CapsNet and performs better than the traditional CapsNet and some other models found in literature on brain tumor detection. The proposed model exhibits better convergence speed and can generalize well on unknown data, hence can serve as an intelligent tool, assisting physicians in diagnosing and administering appropriate

treatments for brain tumors. Fig. 1 depicts the adopted workflow for the proposed work.

Most of the models found in literature performed dataset balancing, data segmentation, data augmentation, and thorough data preparation, before model fitting. This study utilized raw datasets without any data augmentation and data preprocessing to evaluate the proposed model's effectiveness with natural or raw data since data augmentation or segmentation might not be practical in medical emergency situations. Also, this study offers detailed visual representations of image regions that capture the focus of specific parts of our model, evaluation performance on imbalanced datasets using ROC and Precision-Recall (PR) curves, clusters of features at the class capsule layer to assess the model's effectiveness, and the model's transparency and understandability was enhanced by reconstructing input images.

The contributions of this study are:

- 1) A fast, robust, and low-parameterized TTSFM CapsNet, that has efficient feature extraction capabilities is proposed.
- 2) *Separable* convolutions are employed to reduce the size and trainable parameters of the model.
- 3) A comparative analysis was conducted to assess the proposed model with other CapsNet models.
- 4) *The* study presented a comprehensive visual representation of the outputs of layers to help offer notable contributions to the explainable artificial intelligence field.

The study is organized as follows: Related works are presented in Section II. Methodology is presented in Section III. Section IV deals with results and discussion and Section V deals with the conclusion and future works of the study.

## II. RELATED WORKS

Manual diagnosis in the medical field is prone to error, tedious for health professionals, and time-consuming. These limitations led to the employment of algorithms for predicting and detecting radiomic medical conditions. For instance, Gao et al., [11] utilized both 2D and 3D convolutional neural networks (CNNs), the researchers employed these networks to categorize individuals as having tumors, no tumors, or Alzheimer's disease based on CT scans. They were able to attain an accuracy level of 87.6%. A hybrid approach employing CNN and Neutrosophy (NS-CNN), was proposed by Özyurt et al [12]. The approach was used to categorize benign or malignant segmented tumor regions from brain tumor images. The accuracy of the suggested model was 95.62%. Sajjad et al. [13] presented a modified (CNN) based multi-grade system for classifying brain tumor grades. The model's accuracy was 90.67%. In order to solve the classification challenge for brain tumors, Ayadi et al. [14] suggested a new model that makes use of the CNN sequential model. The model has several layers and was designed to categorize MRI brain cancers. The model had a 94.74% accuracy rate. A CNN model was proposed by Badža and Barjaktarovic [15] for classing three distinct forms of brain tumors. The

suggested model, which has a straightforward architecture akin to traditional CNN, accurately classified 96.56% of the brain tumor MRI images in the dataset. Also, Afshar et al. [16], introduced a boosted capsule network (also known as BoostCaps), that makes use of boosting approaches' capacity to accommodate poor learners, by steadily boosting the models. The BootsCaps architecture, according to the results, classified brain tumors with an accuracy of 92.45%. DCNet and DCNet++ were suggested by Phaye et al. [17]. By substituting densely connected convolutions for the typical convolutional layers in the two suggested models, the CapsNet was modified. On Brain Tumor Dataset, the two models were assessed and achieved a validation accuracy of 93.04% and 95.03%, respectively. A capsule network for automatic brain tumor classification that achieves a 92.65% accuracy was proposed by Goceri. This network includes three fully connected layers and utilizes an expectation-maximization (EM)-based dynamic routing algorithm to extract important features from images [18]. In order to increase the focus of CapsNet, Afshar and colleagues suggested an improved CapsNet architecture for classifying brain tumors that incorporates the tumor coarse boundaries as additional inputs. The validation accuracy for the model was 90.89% [19]. According to Adu and friends, an improved CapsNets with several convolutional layers and dilation to preserve image resolution and boost classification accuracy was proposed. The proposed system can guarantee an increase in CapsNets focus by inputting segmented tumor regions within the structure. This model's performance obtained an accuracy of 95.54% [20]. Some researchers presented the BayesCap, a Bayesian CapsNet architecture that can offer both the mean forecasts and entropy as a gauge of uncertainty in forecasting. According to the findings, accuracy can be increased by filtering out uncertain forecasts. The model's maximum accuracy was 73.9% with a CI of: (73.5%, and 74.4%) [21].

All the existing models performed well on the various datasets. But for medical image diagnosis, there is a need for a more robust and efficient model for better diagnosis, hence this study aims to propose an improved, fast, low-parameterized, and robust Capsule Network which incorporates Texton and Separable convolutions for effective feature extraction and better classification of brain tumor diseases. Most of the studies mentioned above performed dataset balancing, data segmentation, data augmentation, and thorough data preparation, before model fitting.

## III. METHODOLOGY

This section presents the methodologies employed to attain our goal of developing a deployable CapsNet that has an effective ability to extract features efficiently with lesser parameters and size on disk. Fig. 1 shows the proposed methods block diagram for the automatic classification of brain tumor types.

### A. Capsule Network

The structure of the baseline CapsNet on which the proposed model is based is found in Fig. 2.



B. Texton Detection

The notion of Texton involves the identification of clusters of shapes within an image that possess a shared characteristic. Julesz further developed this concept [22], placing emphasis on the significance of measuring the distances between texture elements when calculating gradients of Textons. Textures emerge only when neighboring elements are in proximity, and the scale of the elements impacts the surrounding region. Larger elements oriented in a particular direction can slightly impede the initial, instinctive differentiation. Gradients of Texton are only present at the borders of textures, so utilizing a smaller element size, such as 2x2, can enhance the

distinction of textures. The approach of Multi Texton Detection (MTD), utilized to extract information regarding edges and colors, involves the use of six distinct types of Texton ( $T_1, T_2, T_3, T_4, T_5,$  and  $T_6$ ) on a 2x2 grid (shown in Fig. 3) to identify textons. A Texton is generated by the grid when the two shaded pixels share the same value. By systematically shifting the 2x2 block across the image  $C(x, y)$ , textons can be detected in a stepwise manner. If a texton is identified, the original pixel values are preserved; otherwise, the block is disregarded. The resultant image containing textons is represented as  $T(x, y)$  [10], as illustrated in Fig. 4.

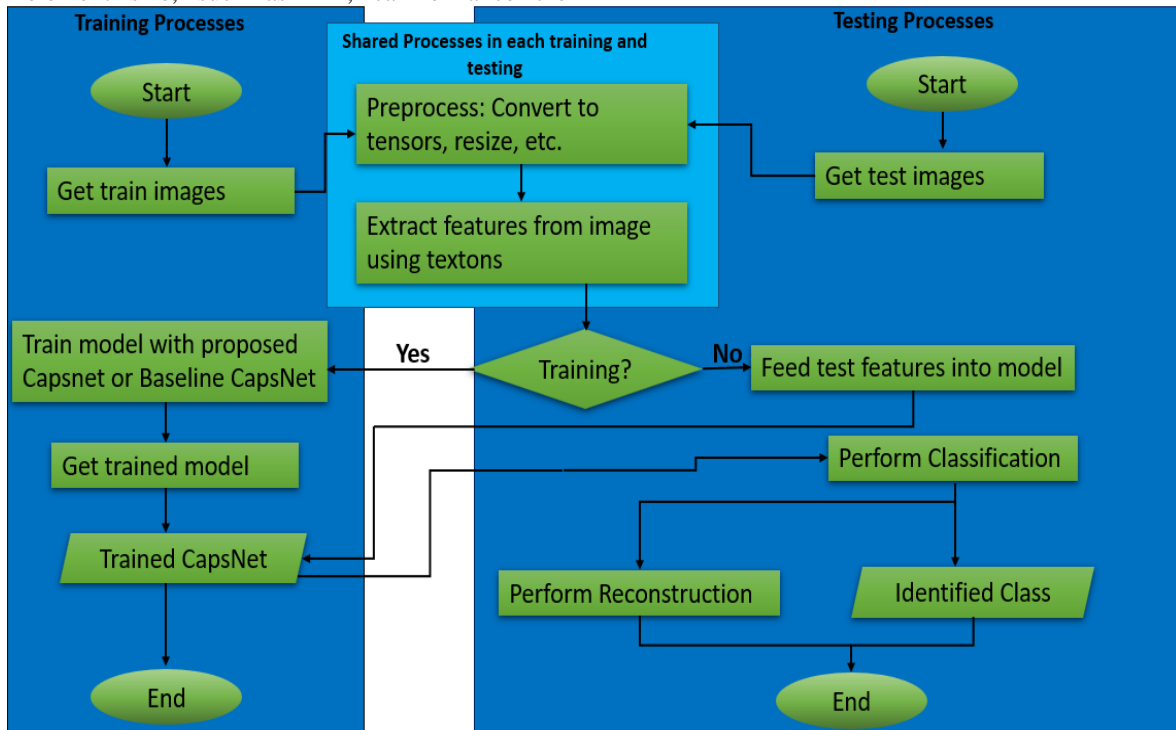


Fig. 1. Workflow diagram of the study.

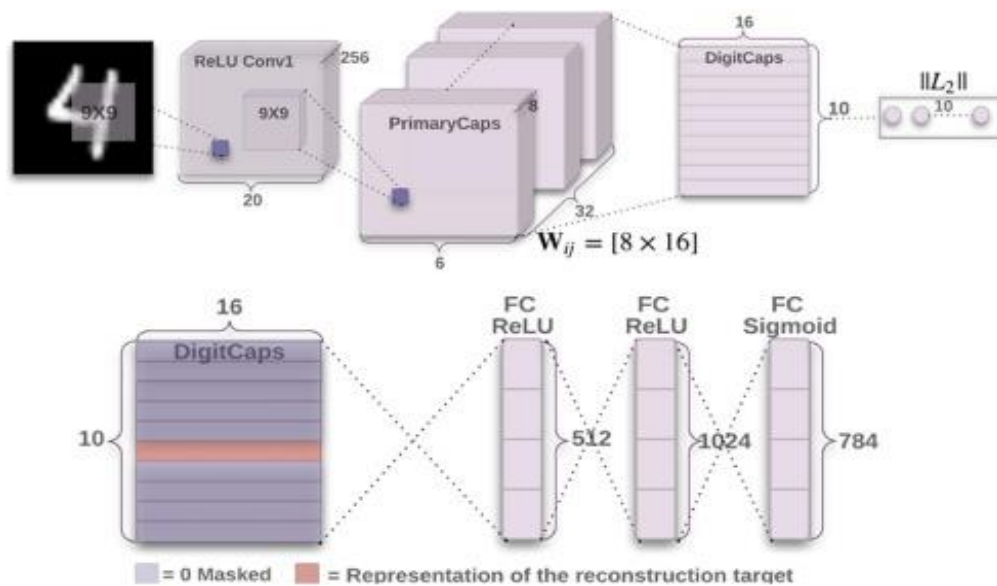


Fig. 2. Architecture of the baseline capsule network model.

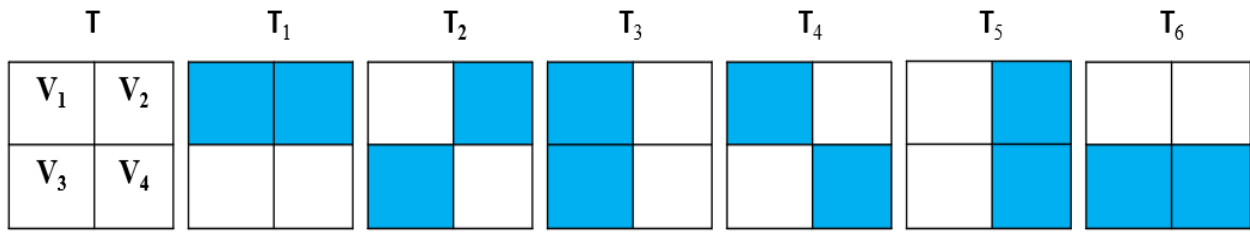


Fig. 3. Six Texton types used in Texton detection process: (T) 2x2 grid.

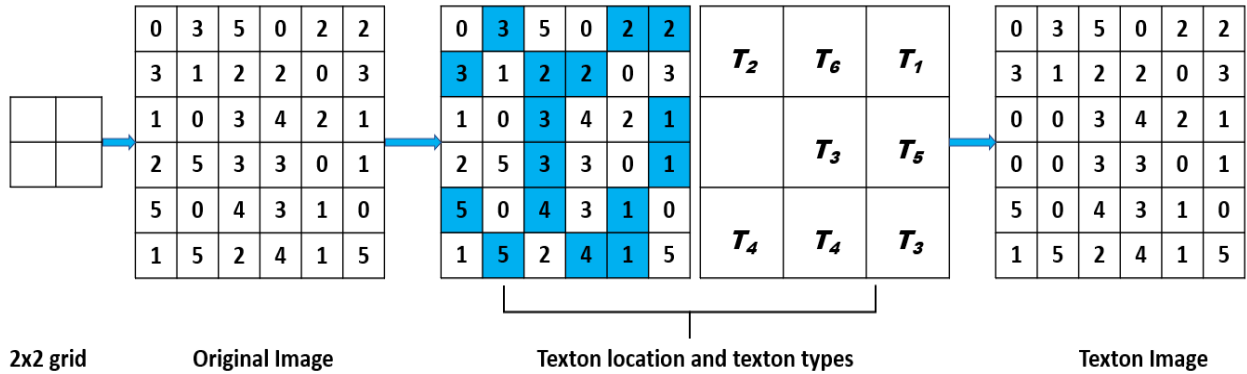


Fig. 4. Illustration of the Texton detection process.

### C. Depthwise Separable Convolution

Two separable convolutions types exist in separable convolutional neural networks. These are depthwise separable convolutions (DSC) and spatial separable convolutions (SSC). DSC is adopted for this study, and can be viewed as grouped convolutions, similar to the concept of 'inception modules' employed in the design of the Xception architecture [23]. It relies on spatial convolution, which operates separately on each input channel. After the spatial convolution, a pointwise convolution (PC) is executed, which involves a standard convolution using  $1 \times 1$  windows. This leads to the creation of a new channel space as a result of projecting the channels calculated during the depthwise convolution (DC). The mathematical expression for depthwise convolution (DC) and pointwise convolution (PC) is presented as follows:

$$PC(W, y)_{(i,j)} = \sum_m W_m \cdot y_{(i,j,m)} \quad (1)$$

$$DC(W, y)_{(i,j)} = \sum_{k,l}^{K,L} W_{(k,l)} \odot y_{(i+k, j+l)} \quad (2)$$

$$DSC(W_p, W_d, y)_{(i,j)} = PC_{(i,j)} \left( W_p, DC_{(i,j)}(W_d, y) \right) \quad (3)$$

$W_p$  and  $W_d$  represent the inputs used for pointwise and depthwise convolutions in the above equations, respectively. The operator  $\odot$  within Eq. (2) pertains to the element-wise multiplication. Consequently, the fundamental idea underpinning depthwise separable convolutions involves splitting the feature extraction process carried out by regular convolutions across a unified "space-cross-channels domain" into two distinct stages: spatial pattern learning and channel fusion. This approach represents a generalization when dealing with convolution operations on 2D or 3D inputs having both relatively independent channels and closely interconnected spatial positions.

### D. Proposed Model

Fig. 5 shows the proposed Texton Tri-alley Separable Feature Merging (TTSFM) CapsNet model that employs Texton, separable convolution, traditional convolutions, max pooling, dropout and reconstruction layers. The Texton layer is used to extract important texture and edge features from the input image. The output features from the Texton layer are processed by three different separable convolutions (each having 32 filters, kernel size of 2x2, depth multiplier of 1, depthwise and pointwise initializers of "ones" and a stride of 1) followed with batch normalization and max pooling, contributing to reduced number of parameters, model size, computational time and complexity of the model, as can be seen at alley\_1\_conv1, alley\_2\_conv1, and alley\_3\_conv1. The feature map from alley\_1\_conv1 serves as input in into alley\_1\_conv2, and the feature maps from alley\_2\_conv1, and alley\_3\_conv1 are merged and serves as input into alley\_2\_conv2, and alley\_3\_conv2. The feature maps from alley\_1\_conv2, alley\_2\_conv2, and alley\_3\_conv2 (all conv2 layers employs 64 filters, kernel size of 3x3, and a stride of 1) are then concatenated and sent as input into a dropout layer, followed with a batch normalization layer. This feature map is sent as input into the primary capsule layer consisting of 32 channels with eight dimensions, a kernel size of 3x3, and a stride of two. Features from this primary capsule are then sent to the TumorCaps layer (by employing dynamic routing algorithm) for classification. This TumorCaps consist of the total classes number in 16D capsules. The output of TumorCaps is directed to the reconstruction layer, which works on rebuilding the characteristics acquired from the TumorCaps. The features are then transferred to the decoder layer within the capsule, which decodes the properties of the entity. This decoder is composed of three layers of fully connected neurons, with counts of 512, 1024, and 3072 respectively. The Texton, max pooling, and convolution layers

help to extract important features from the input images using separable convolutions results in reduced number of parameters, model size, computational time and complexity of the model.

#### E. Dataset Description

Brain Tumor: The dataset comprises 7022 MRI scans of the human brain, which are grouped into four categories: (1) glioma, (2) meningioma, (3) pituitary, and (4) no tumor. To make the dataset more manageable, the images were resized to  $32 \times 32 \times 3$  and redistributed using 70:20:10 leave-out approach. This publicly available dataset can be found at: <https://www.kaggle.com/datasets/masoudnickparvar/brain-tumor-mri-dataset>.

#### F. Experimental Setup

The proposed model was developed and assessed using Keras, Python via Anaconda, and employed the TensorFlow backend on a 64-bit Windows computer. The hardware configuration encompassed an NVIDIA GeForce RTX 2080 SUPER GPU having 8GB of dedicated GPU memory along with 32GB of system RAM. During the training stage, Adam optimizer was employed with a learning rate set to 0.001, and training operations were executed using batches of 100 samples. In order to ensure optimal training progress, the model achieving the highest performance was saved during the training iterations. The evaluation of loss was conducted using the margin loss, represented as  $L_k$  in Eq. (4), with specific details provided below:

$$L_k = T_k \max(0, m^+ - ||v_k||)^2 + \lambda(1 - T_k) \max(0, ||v_k|| - m^-)^2 \quad (4)$$

where,  $T_k$  is 1 when class  $k$  is active and 0 otherwise. Hyper-parameters  $\lambda$ ,  $m$ ,  $m^+$  are set during the learning process.

#### G. Performance Evaluation Measures

The following metrics were employed in this study for the purpose of classification:

Validation Accuracy: Calculates the proportion of accurately classified classes from the total number of classes. The attained overall validation accuracy for the entire set of experiments is reported.

Loss: Evaluates the variance between the model's predictions and the actual labels. This assessment employs the margin loss during testing.

Confusion Matrix: Assist in providing a thorough examination of the tally of correctly and incorrectly categorized images. Factors like True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) are employed to evaluate diverse measurements such as precision, accuracy, specificity, sensitivity (recall), and additional indicators.

Precision (P): The proportion of accurately detected positive instances compared to the overall number of predicted positive instances.

Recall (R) or Sensitivity: the proportion of accurately detected positive instances in relation to the overall count of positive instances within the dataset.

Specificity: The proportion of negative instances have been accurately recognized in relation to the overall count of negative instances present in the dataset.

F1-Score: The Mean that combines precision and recall in a harmonic manner.

Area under the curve (AUC): The model's performance is assessed on datasets where classes are imbalanced or unevenly distributed by creating Receiver Operating Characteristic (ROC) and precision-recall (PR) curves [24]-[25]. Higher AUC values are favored compared to their smaller equivalents.

Clustering: We utilize t-distributed stochastic neighbor embedding to acquire and examine the clusters within the class capsule layer of the models.

## IV. RESULTS AND DISCUSSIONS

In this section, we present the outcomes of our experiments and demonstrate the favorable performance of the model when tested on the brain tumor dataset in comparison with the baseline Capsule Network [5]. To bolster confidence and ensure the reliability of the model's outcomes, we employed and meticulously executed various evaluation methods. These techniques included evaluating metrics such as classification accuracy and loss, specificity, sensitivity, precision, F1-Score, number of parameters, Area Under the Curve (AUC) for both the Receiver Operating Characteristic Curve (ROC) and Precision-Recall (PR) curves. We also trained a traditional capsule network [5] using the same dataset and compared its results with our model's performance using the aforementioned metrics.

### A. Performance Evaluation

Graphs presented in Fig. 6 illustrate the accuracy and loss trends for both CapsNet models: the proposed and the baseline model. The accuracy and loss graphs during training and validation reveals that the proposed model outperform the baseline CapsNet model, exhibiting superior and consistent accuracy with quicker convergence. It is important to highlight that while accuracy is widely used to assess classification algorithms, it may not be suitable for evaluating medical images due to their small size and significant class imbalance [26]. Despite its limitations, accuracy can offer an overview of overall system.

Fig. 7 displays the ROC and PR curves for both the proposed and baseline models. Analyzing the data from these curves, it becomes evident that the performance of the proposed model surpasses that of the baseline model. This shows that the proposed model performs better on small and imbalanced datasets like medical images [27] than the baseline model.

Fig. 8 depicts confusion matrices illustrating the accurate and erroneous image identifications. The results presented in Table I highlight that the proposed model outperformed the CapsNet baseline by exhibiting fewer misclassifications.

Hence resulted in better per class accuracy, specificity, sensitivity, precision and F1-Score for each class, as compared to the baseline Capsule Network model.

### B. Ablation Study

Conducting ablation experiments involves analyzing the elements of the model that significantly influenced its performance [28][29]. The model's layers are systematically removed in succession to assess their effects on the overall model's performance. As depicted in Table II, the model's performance shows a significant improvement through the integration of the Texton and max-pooling layers.

### C. Number of Parameters and Size on Disk

A number of models found in literature expand their width and depth in order to enhance their performance on complex images. This causes a surge in the number of parameters. For example, ResNet50 [30], AlexNet [31], and VGG16 [32], among others, generate parameter counts of 23 million, 60 million, and 138 million respectively. The intricacy of a model directly correlates with the parameters it generates, resulting

in a substantial computational load that strains the resources of a system. Consequently, this poses a constraint on the feasibility of deploying such models on devices with limited memory, such as mobile phones. The comparison of the parameters of the models as well as size on disk is found in Table III. It can be seen that the size of the model on disk is small and less parameters were generated by the proposed model. This makes the proposed model suitable for deployment on mobile devices.

### D. Model Interpretability

The inner workings of deep learning models are often labeled as black boxes. In order to rely on and employ these models for important functions, such as in the field of healthcare, it is essential that both the operations within the models and the results they produce are explainable. Through the utilization of saliency maps, mathematical models, activation maps, and similar techniques, explainable neural networks [33][34] and model interpretability [29] approaches aid in revealing insights into the operations occurring within the inner layers of deep learning models.

TABLE I. PERFORMANCE METRICS ON THE BRAIN TUMOR DATASET FOR THE PROPOSED AND BASELINE CAPSNET MODELS

Model (Dataset)	Class	TP	FP	TN	FN	Precision	Sensitivity	Specificity	Accuracy	F1-Score	Data Size
Baseline (Brain Tumor)	0	271	16	995	29	0.9443	0.9033	0.9842	96.57%	0.9235	300
	1	283	30	975	23	0.9042	0.9248	0.9702	95.96%	0.9144	306
	2	296	6	1005	4	0.9801	0.9867	0.9941	99.24%	0.9834	300
	3	405	4	902	0	0.9902	1	0.9956	99.70%	0.9951	405
Proposed (Brain Tumor)	0	287	10	1001	13	0.9663	0.9569	0.9872	98.25%	0.9616	300
	1	289	12	993	17	0.9601	0.9444	0.9832	97.79%	0.9522	306
	2	299	9	1002	1	0.9708	0.9967	0.9990	99.24%	0.9836	300
	3	405	0	906	0	1	1	1	100%	1	405

TABLE II. ABLATION STUDY RESULTS

Layers	Validation accuracy %
-texton	95.04
-alley_1_conv1	96.49
-alley_2_conv1	96.11
-alley_3_conv1	96.11
-alley_1_conv2	97.48
-alley_2_conv1	97.41
-alley_3_conv1	97.41
+ all layers	97.64

TABLE III. COMPARISON OF PARAMETERS OF MODELS AND SIZE ON DISK

Model	Trainable Parameters	Non-Trainable Parameters	Size on disk
Baseline CapsNet model	10,127,104	0	38.6MB
Proposed model	4,834,532	960	18.5MB

### E. Visualization of Activation Maps and Clusters

Here, comparison of activation maps and clusters from the proposed and baseline models are done. This help to know the model that extract more important features from input images. Proposed model features from the Texton layer, extracted by

one of the first separable convolutions is shown in Fig. 9. 1<sup>st</sup> row image 1 and baseline model features extracted by the convolution layer is shown in Fig. 9. In 2<sup>nd</sup> row image 1, insufficient features were extracted. This exhibit that, the baseline convolution layer alone is not enough to extract important features. This inability of the convolution layer of the baseline model affected the primary capsule layer since it did not extract more important necessary to make differentiation between capsules, whereas the proposed model convolution layer extracted better edge and textural features from the Texton layer, hence its primary capsule produced better activation maps as seen in Fig. 9 1<sup>st</sup> row image 2 than the baseline PC activation maps seen in Fig. 9 2<sup>nd</sup> row image 2. These visualizations of layers help to improve the understandability of the inner workings of the black box models and contributes to explainable Artificial Intelligence [35][36][37].

The technique of t-distributed stochastic neighbor embedding (tsne) [38] [39], was employed to visually represent the distinctness of clusters formed within the class capsule layer of the models. The suggested model displays noticeable groupings in contrast to the clusters formed by the baseline model. While a few outliers are evident in both the suggested and baseline model clusters, the outliers in the suggested model remain relatively close to their respective clusters. This highlights the effective discriminatory capability

of the suggested model in comparison to the baseline model as it can be seen in Fig. 10.

F. Prediction and Reconstruction

The process of determining the likely class of an input image and determining if there is a strong likelihood for that categorization is achieved through the application of a

reconstruction method. In light of this, this research showcases reconstructed images of Brain Tumor using the decoder network for both models. The images generated by the proposed model exhibit slightly improved visual quality and demonstrate higher class identification and emphatic probabilities per class compared to the images produced by the baseline model, as observed in Fig. 11.

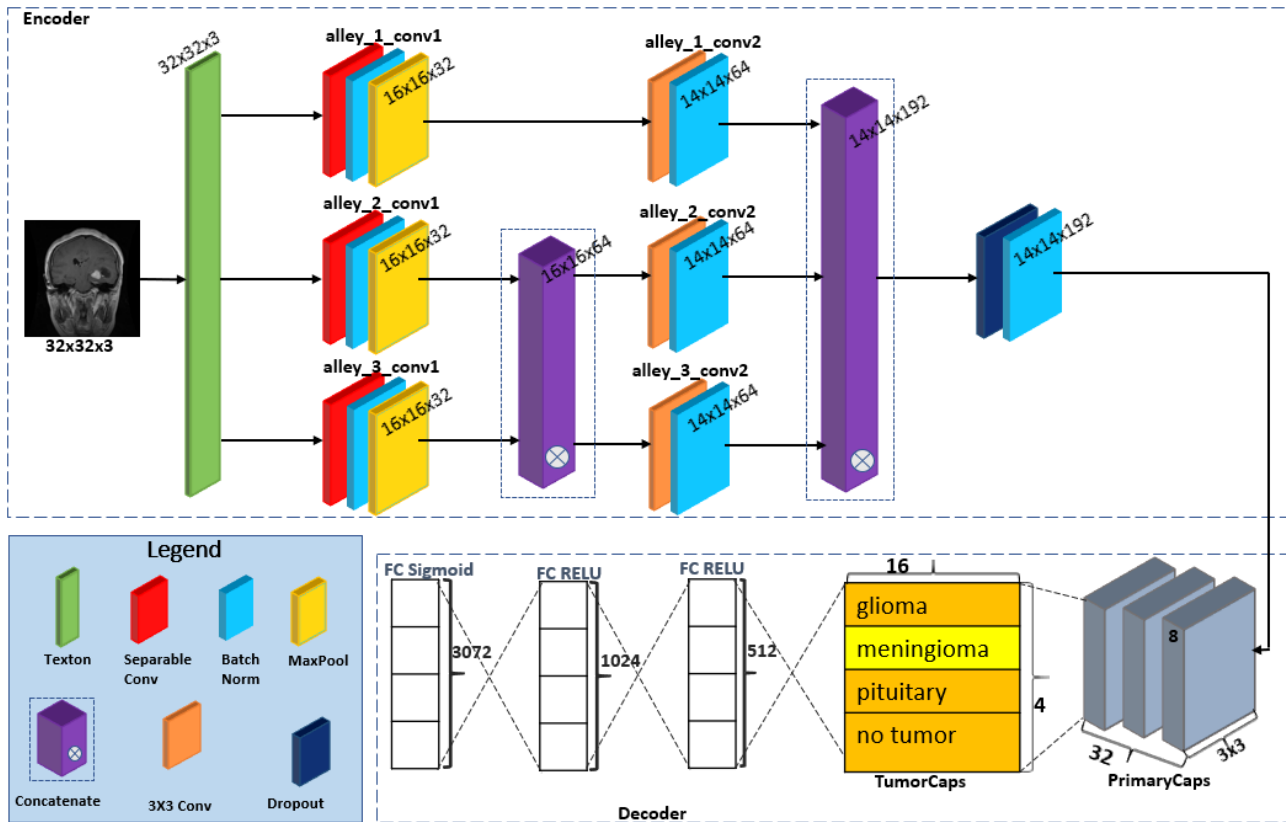


Fig. 5. Architecture of the proposed CapsNet model.

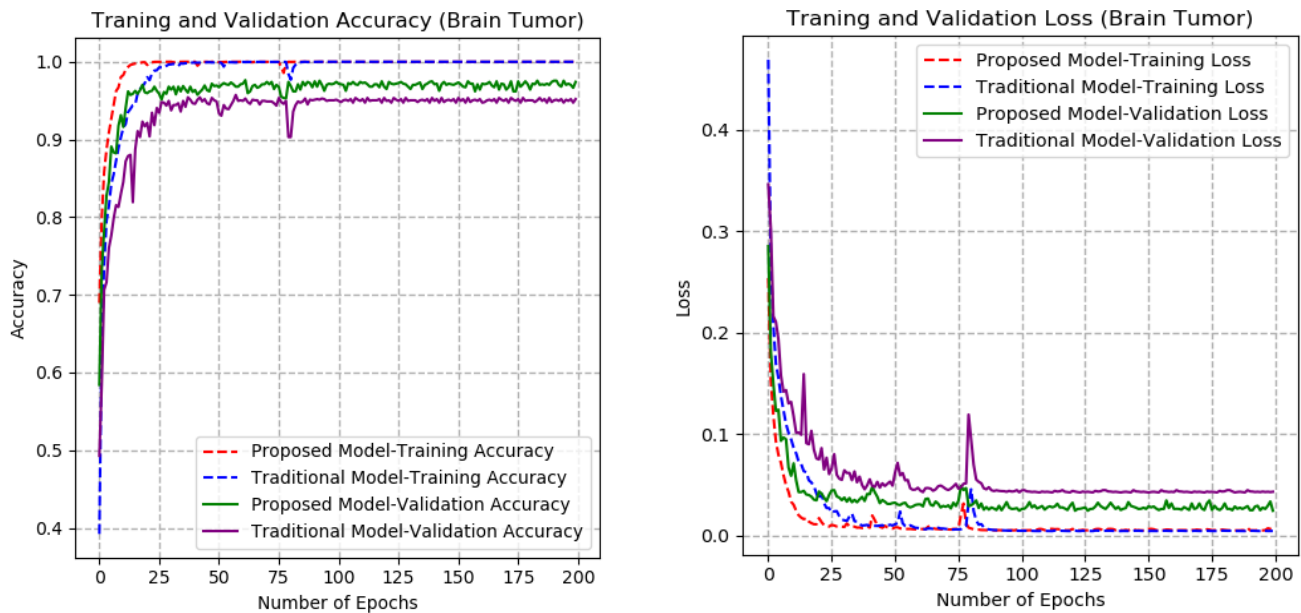


Fig. 6. Accuracy and Loss graphs of the proposed and baseline CapsNet models.

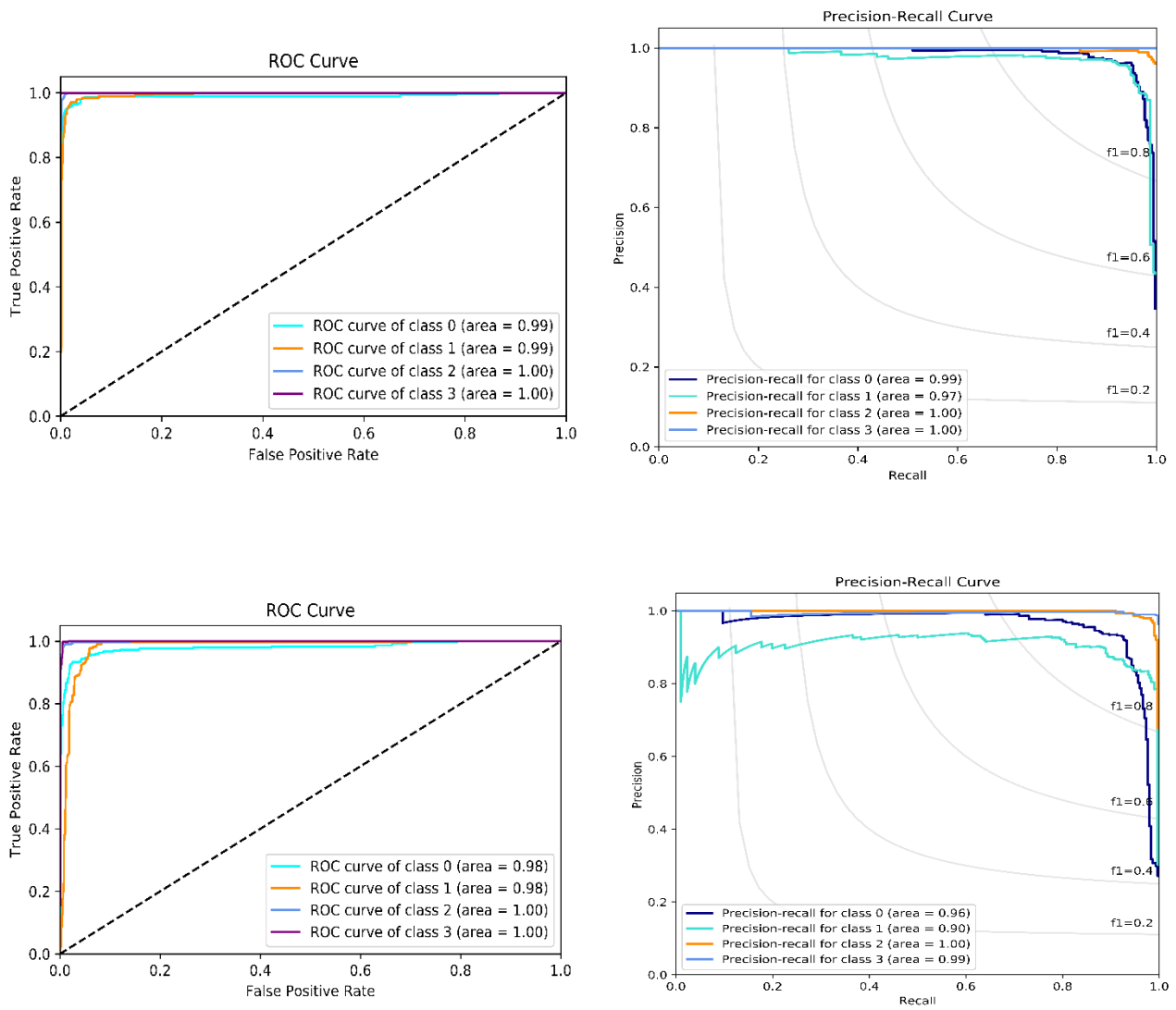


Fig. 7. ROC and PR curves for the (1<sup>st</sup> row) proposed and (2<sup>nd</sup> row) baseline models.

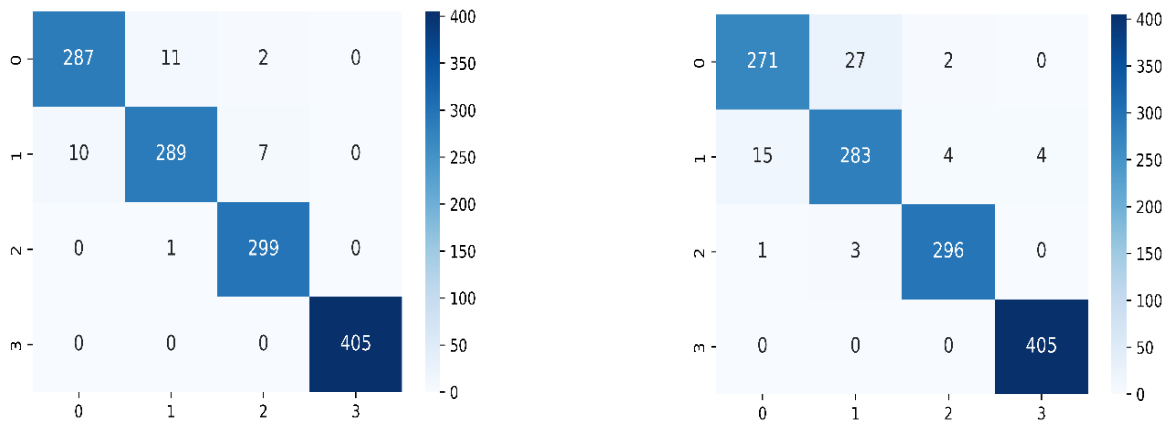


Fig. 8. Confusion Matrices of (left) proposed and (right) baseline models.

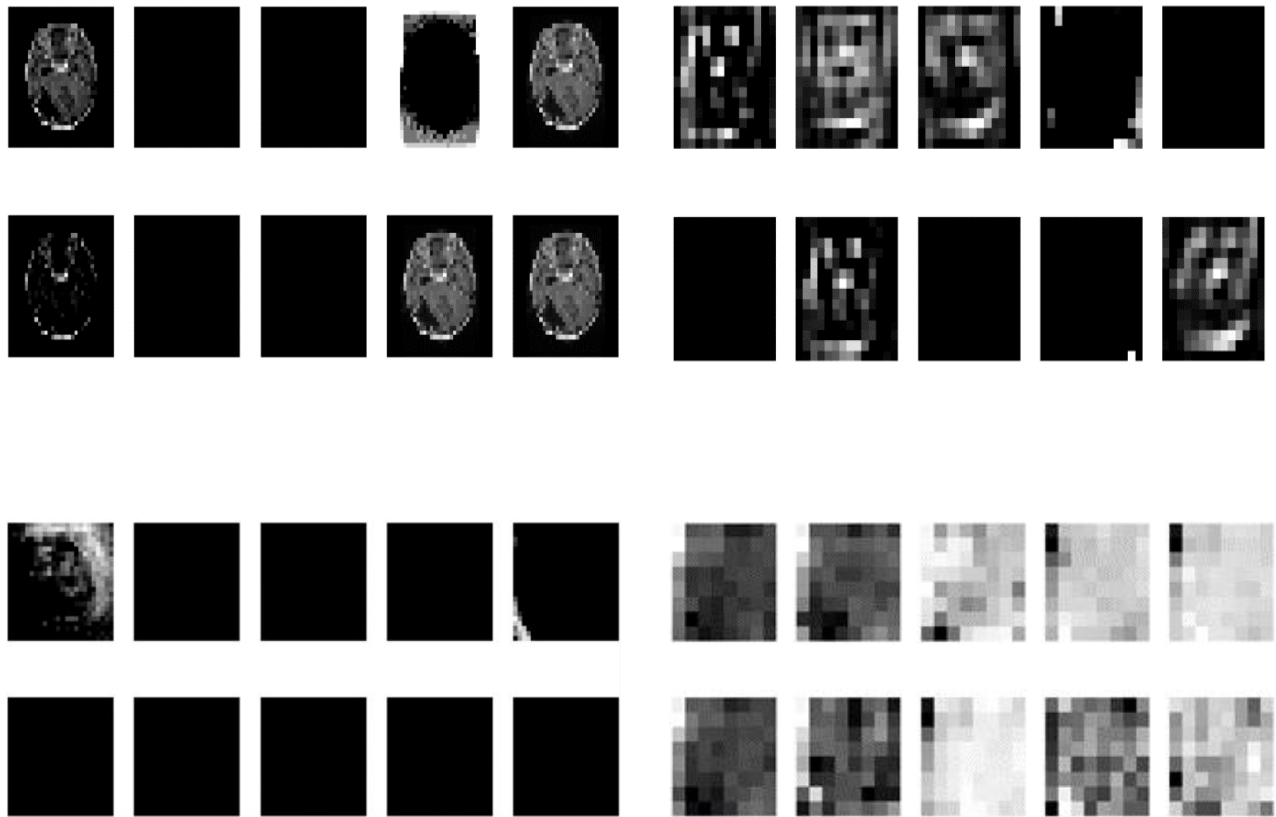


Fig. 9. Activation maps from the proposed and baseline convolution and primary capsule layers.

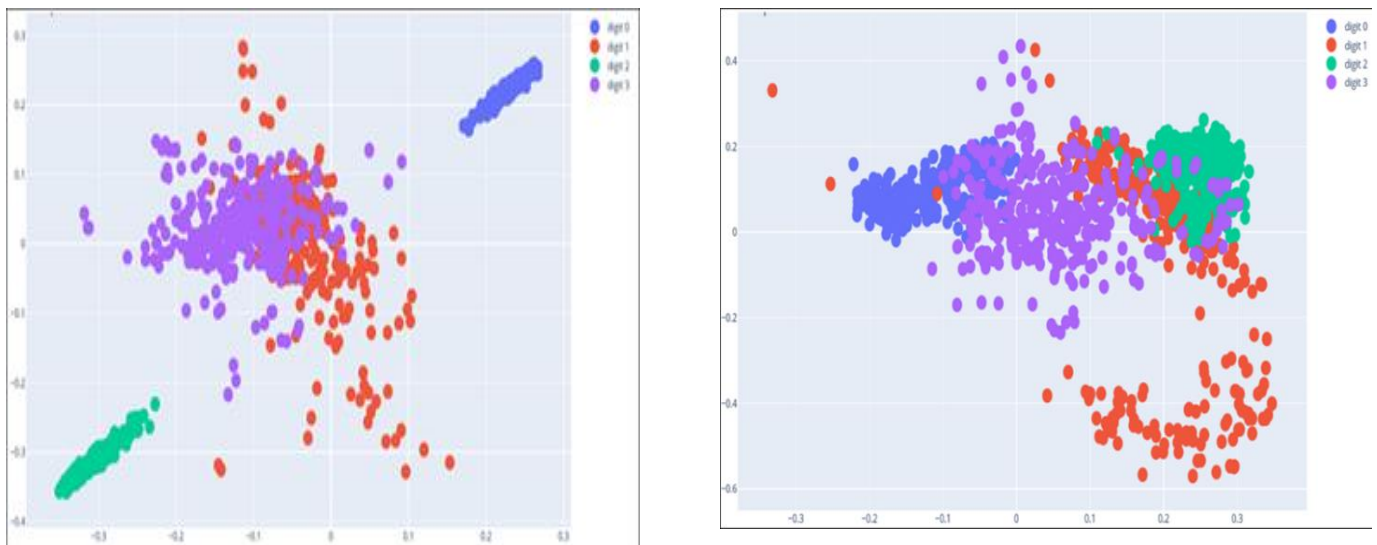


Fig. 10. Clusters obtained by the (left) proposed and (right) baseline models class capsules.

### G. Comparison of Results

To demonstrate the effectiveness of our novel approach, we conducted a performance comparison between our model and cutting-edge models on brain tumor datasets. Our modifications primarily center around the structure of capsule network, specifically focusing on dynamic routing. Although our main emphasis was on dynamic routing, we extended our

investigation to encompass multiple routing techniques. The outcomes, as detailed in Table IV, indicate that our model's performance matches that of the current state-of-the-art capsule network models. The commendable performance achieved by our proposed model in medical image diagnosis can be attributed to its adeptness in extracting pertinent information from diverse images, which contributes to its capability in achieving accurate results.

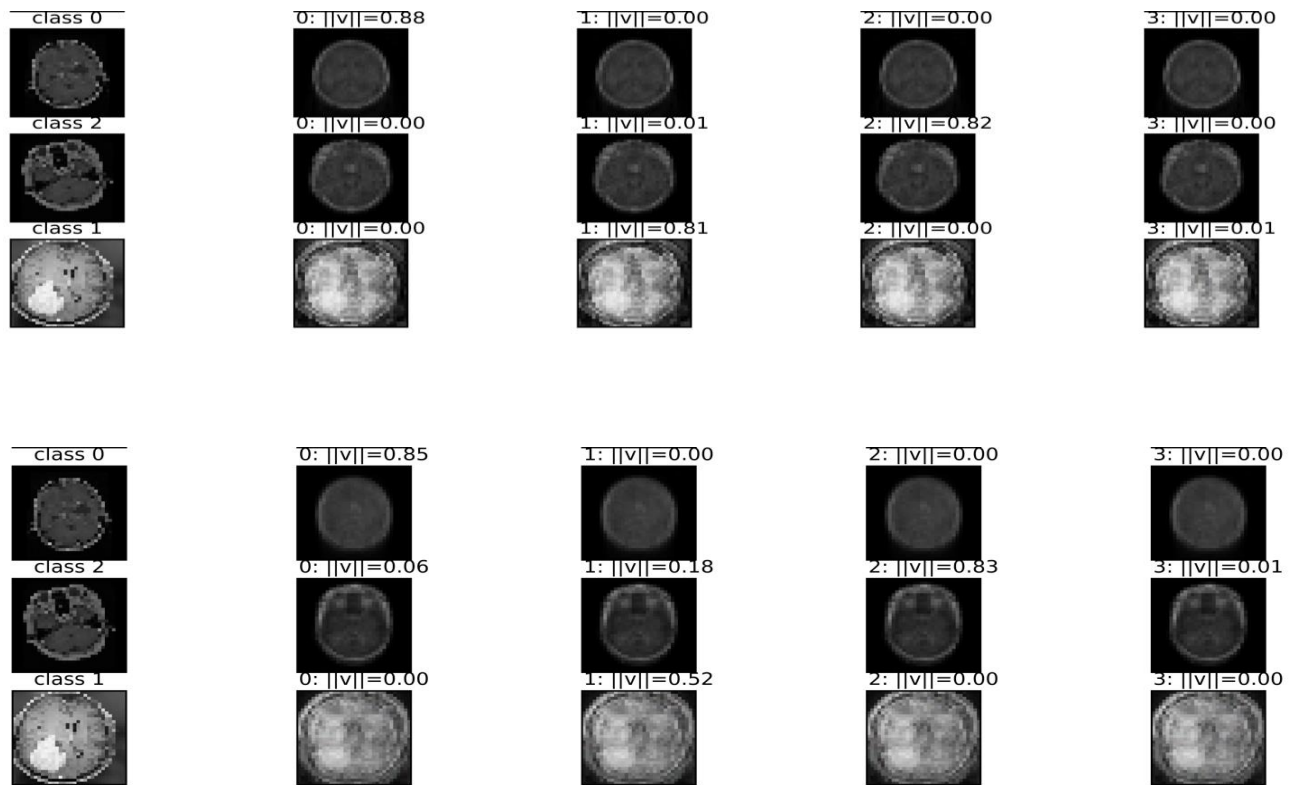


Fig. 11. Reconstructed images (top) proposed and (down) baseline model.

TABLE IV. PROPOSED MODEL AND PREVIOUS WORKS COMPARISON ON BRAIN TUMOR DATASET

CapsNet Methods	Validation accuracy (%)
Baseline [5]	95.73
BoostCaps [16]	92.45
DCNet and DCNet++ [17]	93.04 and 95.03
MLAF-CapsNet[40]	93.40 and 96.60
Vimal Kurup et al.[41]	92.60
Afshar et al. [19]	90.89
Dilated CapsNet. [20]	95.54
BayesCaps[21].	73.9
<b>Proposed Model</b>	<b>97.64</b>

## V. CONCLUSION AND FUTURE WORKS

This study introduced a novel architecture that utilizes less time to train with less parameters, small size on disk, and proficient feature extraction capabilities, named Texton Tri-alley Separable Feature Merging (TTSEFM) CapsNet, utilizing a capsule network approach, aimed at the detection of brain tumors. Texton layer helps to extract important features from input image and the separable convolutions coupled with the use of less filters and kernel sizes resulted in using less amount of time for training, small size on disk, and a smaller number of trainable parameters. These components and properties lead to the appreciable performance of the proposed model, making the model deployable on devices with lower memory like mobile devices. We went on to enhance the model's interpretability and practical usability by conducting

thorough analyses, including extensive visualization of layer activation maps, examination of feature clusters, and performing ablation study.

In future, our focus will be on improving the performance of the model, and conducting in-depth experiments using medical datasets to advance the field of explainable artificial intelligence (XAI). Our objective is to remove all uncertainties from the outcomes of the models, ensuring that both professionals and other users can trust the reliable application of these models in disease diagnosis.

## REFERENCES

- [1] D. N. Louis et al., "The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary," *Acta Neuropathol.*, vol. 131, no. 6, pp. 803–820, Jun. 2016, doi: 10.1007/s00401-016-1545-1.
- [2] K. D. Miller et al., "Cancer treatment and survivorship statistics, 2022," *CA. Cancer J. Clin.*, vol. 72, no. 5, pp. 409–436, 2022, doi: 10.3322/caac.21731.
- [3] A. F. M. SAIF, C. SHAHNAZ, W.-P. ZHU, and M. O. AHMAD, "Abnormality Detection in Musculoskeletal Radiographs Using Capsule Network," *IEEE Access*, vol. 7, pp. 81494–81503, 2019, doi: 10.1109/ACCESS.2019.2923008.
- [4] N. Tajbakhsh et al., "Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1299–1312, 2016, doi: 10.1109/TMI.2016.2535302.
- [5] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic Routing Between Capsules," *Appl. Biosaf.*, vol. 22, no. 4, pp. 185–186, 2017, doi: 10.1177/1535676017742133.
- [6] X. Zhang et al., "Real-time gastric polyp detection using convolutional neural networks," *PLoS One*, vol. 14, no. 3, pp. 1–16, 2019, doi: 10.1371/journal.pone.0214133.t005.



- [7] A. Singh, S. Sengupta, and V. Lakshminarayanan, "Explainable deep learning models in medical image analysis," *J. Imaging*, vol. 6, no. 6, pp. 1–19, 2020, doi: 10.3390/JIMAGING6060052.
- [8] G. E. Hinton, A. Krizhevsky, and S. D. Wang, "Transforming auto-encoders," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 6791 LNCS, no. PART 1, pp. 44–51, 2011, doi: 10.1007/978-3-642-21735-7\_6.
- [9] M. Kwabena Patrick, A. Felix Adekoya, A. Abra Mighty, and B. Y. Edward, "Capsule Networks – A survey," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 1, pp. 1295–1310, 2022, doi: 10.1016/j.jksuci.2019.09.014.
- [10] G.-H. Liu, L. Zhang, Y.-K. Hou, Z.-Y. Li, and J.-Y. Yang, "Image retrieval based on multi-texton histogram," *Pattern Recognit.*, vol. 43, no. 7, pp. 2380–2389, Jul. 2010, doi: 10.1016/j.patcog.2010.02.012.
- [11] X. W. Gao, R. Hui, and Z. Tian, "Classification of CT brain images based on deep learning networks," *Comput. Methods Programs Biomed.*, vol. 138, pp. 49–56, 2017, doi: 10.1016/j.cmpb.2016.10.007.
- [12] F. Özyurt, E. Sert, E. Avci, and E. Dogantekin, "Brain tumor detection based on Convolutional Neural Network with neutrosophic expert maximum fuzzy sure entropy," *Meas.* 2019, vol. 147, 2019, doi: 10.1016/j.measurement.2019.07.058.
- [13] M. Sajjad, S. Khan, K. Muhammad, W. Wu, A. Ullah, and S. Wook, "Multi-grade brain tumor classification using deep CNN with extensive data augmentation," *J. Comput. Sci.*, vol. 30, pp. 174–182, 2019, doi: 10.1016/j.jocs.2018.12.003.
- [14] W. Ayadi, W. Elhamzi, I. Charfi, and M. Atri, "Deep CNN for Brain Tumor Classification," *Neural Process. Lett.*, vol. 53, no. 1, pp. 671–700, 2021, doi: 10.1007/s11063-020-10398-2.
- [15] M. M. Badža and M. c Barjaktarovic, "Classification of Brain Tumors from MRI Images Using a Convolutional Neural Network," *Appl. Sci.*, 2020.
- [16] P. Afshar, K. N. Plataniotis, and A. Mohammadi, "BoostCaps: A Boosted Capsule Network for Brain Tumor Classification," 2020 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., pp. 1075–1079, 2020.
- [17] S. S. R. Phayre, A. Sikka, A. Dhall, and D. Bathula, "Dense and Diverse Capsule Networks: Making the Capsules Learn Better," *Comput. Vis. Pattern Recognit.*, pp. 1–11, May 2018, [Online]. Available: <http://arxiv.org/abs/1805.04001>
- [18] E. Goceri, "CapsNet topology to classify tumours from brain images and comparative evaluation," *IET Image Process.*, vol. 14, no. 5, pp. 882–889, 2020, doi: 10.1049/iet-ipr.2019.0312.
- [19] P. Afshar, A. Mohammadi, and K. N. Plataniotis, "Brain tumor type classification via capsule networks," 2018 25th IEEE Int. Conf. Image Process., 2018.
- [20] K. Adu, Y. Yu, J. Cai, and N. Tashi, "Dilated Capsule Network for Brain Tumor Type Classification Via MRI Segmented Tumor Region," *Proceeding IEEE Int. Conf. Robot. Biomimetics Dali, China*, December 2019, no. December, pp. 942–947, 2019.
- [21] P. Afshar, A. Mohammadi, and K. N. Plataniotis, "BayesCap: A Bayesian Approach to Brain Tumor," *IEEE Signal Process. Lett.*, vol. 27, pp. 2024–2028, 2020.
- [22] B. Julesz, "Texton Gradients: The Texton Theory Revisited," vol. 251, pp. 245–251, 1986.
- [23] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 1251–1258, 2017, doi: 10.4271/2014-01-0975.
- [24] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size," *Comput. Vis. Pattern Recognit.*, pp. 1–13, 2016, [Online]. Available: <http://arxiv.org/abs/1602.07360>
- [25] C. Zhang, S. Bengio, M. Hardt, B. Recht, and O. Vinyals, "Understanding deep learning (still) requires rethinking generalization," *Commun. ACM*, vol. 64, no. 3, pp. 107–115, 2021, doi: 10.1145/3446776.
- [26] F. Provost, T. Fawcett, and R. Kohavi, "The case against accuracy estimation for comparing induction algorithms," *Int. Conf. Mach. Learn.*, p. 445, 1998.
- [27] P. Singla and P. Domingos, "Discriminative training of Markov logic networks," *Proc. Natl. Conf. Artif. Intell.*, vol. 2, pp. 868–873, 2005.
- [28] R. Meyes, M. Lu, and T. Meisen, "Ablation Studies to Uncover Structure of Learned Representations in Artificial Neural Networks," *Proc. Int. Conf. Artif. Intell. (pp. 185-191). Steer. Comm. World Congr. Comput. Sci. Comput. Eng. Appl. Comput. (WorldComp)*, pp. 185–191, 2019.
- [29] R. Meyes, M. Lu, C. W. De Puiseau, and T. Meisen, "Ablation Studies in Artificial Neural Networks," *Comput. Vis. Pattern Recognit.*, pp. 1–19, 2019.
- [30] K. He, X. Zhang, S. Ren, and J. Sun, "Deeper neural networks are more difficult to train," *Comput. Vis. Pattern Recognit.*, vol. 37, no. 50, pp. 1951–1954, Dec. 2016, [Online]. Available: <https://onlinelibrary.wiley.com/doi/10.1002/chin.200650130>
- [31] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Proc. Adv. Neural Inf. Process. Syst. 25 (NIPS 2012)*, pp. 1–1432, 2012, doi: 10.1201/9781420010749.
- [32] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc., pp. 1–14, 2015.
- [33] G. Ras, N. Xie, M. Van Gerven, and D. Doran, "Explainable Deep Learning: A Field Guide for the Uninitiated," *J. Artif. Intell. Res.*, vol. 73, pp. 329–397, Jan. 2022, doi: 10.1613/jair.1.13200.
- [34] P. Angelov and E. Soares, "Towards explainable deep neural networks (xDNN)," *Neural Networks*, vol. 130, pp. 185–194, Oct. 2020, doi: 10.1016/j.neunet.2020.07.010.
- [35] W. Samek and K.-R. Müller, "Towards Explainable Artificial Intelligence," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11700 LNCS, 2019, pp. 5–22. doi: 10.1007/978-3-030-28954-6\_1.
- [36] A. Shahroudjedjad, P. Afshar, K. N. Plataniotis, and A. Mohammadi, "IMPROVED EXPLAINABILITY OF CAPSULE NETWORKS: RELEVANCE PATH BY AGREEMENT," in 2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Nov. 2018, pp. 549–553. doi: 10.1109/GlobalSIP.2018.8646474.
- [37] D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf, and G.-Z. Yang, "XAI—Explainable artificial intelligence," *Sci. Robot.*, vol. 4, no. 37, Dec. 2019, doi: 10.1126/scirobotics.aay7120.
- [38] P. Hajibabae, F. Pourkamali-Anaraki, and M. A. Hariri-Ardebili, "An Empirical Evaluation of the t-SNE Algorithm for Data Visualization in Structural Engineering," in 2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA), Dec. 2021, pp. 1674–1680. doi: 10.1109/ICMLA52953.2021.00267.
- [39] S. Arora, W. Hu, and P. K. Kothari, "An Analysis of the t-SNE Algorithm for Data Visualization," vol. 75, no. 2008, pp. 1–8, Mar. 2018, [Online]. Available: <http://arxiv.org/abs/1803.01768>
- [40] K. Adu, Y. Yua, J. Caia, P. K. Mensah, and K. Owusu-Agyemang, "MLAF-CapsNet: Multi-lane atrous feature fusion capsule network with contrast limited adaptive histogram equalization for brain tumor classification from MRI images," *J. Intell. Fuzzy Syst.*, 2021, doi: 10.3233/JIFS-202261.
- [41] R. Vimal Kurup, V. Sowmya, and K. P. Soman, "Effect of Data Pre-processing on Brain Tumor Classification Using Capsulenet," *ICICCT 2019 – Syst. Reliab. Qual. Control. Safety, Maint. Manag.*, pp. 110–119, 2020, doi: 10.1007/978-981-13-8461-5\_13.

# Unraveling Ransomware: Detecting Threats with Advanced Machine Learning Algorithms

Karam Hammadeh, M. Kavitha

Department of Computer Science and Engineering,  
Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, India

**Abstract**—In our contemporary world, the pervasive influence of information technology, computer engineering, and the Internet has undeniably catalyzed innovation, fostering unparalleled economic growth and revolutionizing education. This technological juggernaut, however, has unwittingly ushered in a parallel era of new criminal frontiers, a magnet for hackers and cybercriminals. These malevolent actors exploit the vast expanse of electronic devices and interconnected networks to perpetrate an array of cybercrimes, and among these insidious digital threats, ransomware reigns supreme. Ransomware, characterized by its ominous ability to encrypt victims' data and extort payment for its release, stands as a dire menace to individuals and organizations alike. Operating with stealth and propagating with alarming alacrity through digital networks, ransomware has emerged as a formidable adversary in the digital age. This research paper focuses on the evolving stages of ransomware, driven by cutting-edge technologies, and proposes essential methods and ideas to detect and combat this menace. The proposed methodology, anchored in Cuckoo Sandbox, PE file feature extraction, and YARA rules, orchestrates three crucial phases: data collection, feature selection, and data preprocessing, all harmonizing to strengthen our defense against this concealed cyber menace. This paper contributes to the development of effective solutions for detecting and mitigating this hidden and insidious cyber threat. This work involves the application of multiple machine learning algorithms, including LSTM, which achieves an impressive accuracy of 99% in identifying ransomware attacks.

**Keywords**—Ransomware; cuckoo sandbox; PEFile; YARA rules; machine learning; LSTM

## I. INTRODUCTION

Cybersecurity has emerged as a crucial domain within information technology and computer engineering due to the rapid advancement and widespread adoption of technologies in our daily lives. As technology continues to evolve, it brings both positive and negative impacts, making it essential to protect data, preserve individual privacy, and safeguard innovation and intellectual property [1]. The primary objective of cybersecurity is to combat cybercrimes, which have significantly increased since the beginning of the 21st century.

Cybercrime is recognized as one of the most damaging and costly forms of criminal activity. Hackers and criminals exploit the power of networks, programming, and computers to steal valuable data, gain unauthorized access to bank accounts, and mass significant financial gains. Their illegal activities often remain hidden, making it challenging to trace the perpetrators and understand the extent of their actions. There are numerous

methods through which individuals can become involved in cybercrime [2]. One such approach is the development and dissemination of malicious code, such as malware and ransomware, which can wreak havoc on computer systems and networks. Another technique employed by cybercriminals is launching Distributed Denial-of-Service (DDoS) attacks, which dominate servers and disrupt their ability to provide services.

Malware, or malicious software, is a wide-open problem that is difficult to solve in computer science [3]. It is a term used to describe various forms of harmful software designed to compromise computer systems, steal data, or disrupt normal operations. There are different types of malwares, including viruses, worms, Trojan horses, and ransomware [4,5]. Malware aims to change the behavior of either the operating system kernel or some security-sensitive applications without the user's consent, and all services of the system will therefore be undocumented. Some countries are developing this kind of malicious application in central intelligence for spying purposes; individuals or teams are also developing it for hijacking or showing their talent [6, 7].

Ransomware is a serious threat that poses a risk to individuals and organizations. It develops rapidly because it uses newer techniques like RSA and C&C servers. These techniques are difficult to be analyzed and to be detected (binary files, payload) [8, 9]. The first ransomware attack occurred in 1989. It was a trojan called PC Cyborg/AIDS, which was created to hide the folders and encrypt the names of all the files. It targeted the files associated with the ADIS conference, and restoration is achievable provided the filename and extension encryption tables are discovered [10].

At the beginning of this decade, ransomware used new techniques. In 2013, RSA 2048-bit was the main characteristic that was implemented with the public key algorithm for ransomware. After that, it could be used as a command and control (C&C) server to communicate through the Tor network, and it became able to target only specific types of file extensions [9, 11].

Encryption methods in ransomware have been significantly developed. For example, in 2015, most ransomware families were using the default configuration of AES file encryption and getting payment via Bitcoin. Later in 2017, the encryption became hybrid, using the AES algorithm to encrypt the files and RSA to encrypt the AES key [11]. According to [8, 12, 13], ransomware has been separated by researchers into many major types according to how it works methodologically and

what kinds of effects that will cause. First, Crypto-Ransomware, Once the ransomware infects the target's device, the virus stealthily detects and encrypts the victim's important files, and the user will not be able to access, use, or get data until he pays the ransom. Second, Locker-Ransomware, which doesn't encrypt the data, has turned its focus to blocking access to the user's equipment and information by disabling the user interface. Third, MBR-Ransomware works to encrypt the Master Boot Record table on the victim's PC, which means the files will be safe and not affected by encryption, but the system will not recognize the location of files and their mac-Ransomware and Mobile-Ransomware.

Some of the researchers [8, 14, 15] also found that there are three types of ransomware according to the style of encryption algorithm that has been used (see Fig. 1). These types are:

- **Symmetric Ransomware:** It uses symmetrical encryption algorithms like AES and DES to encrypt the victim's data; in this type, the encryption and decryption use the same key.
- **Asymmetric Ransomware:** It uses an asymmetrical encryption algorithm, and ransomware is embedded with a public key to encrypt the victim's data, or it downloads during connection with a command and control (C&C) server, but the private key is saved only with the attacker, which is impossible to get without paying.
- **Hybrid Ransomware:** It uses symmetrical encryption to encrypt the victim's files, but the symmetric key will be encrypted by an asymmetric encryption algorithm. This technique takes advantage of both symmetric and asymmetric encryption.

The 2022 update of the Verizon Data Breach Investigations Report (DBIR), as shown in Fig. 2, reveals a concerning trend in the rise of ransomware attacks. According to the report, the number of ransomware incidents surged by 13% between 2020 and 2021, surpassing the combined increase of the previous five years. This significant escalation in ransomware incidents highlights the growing threat posed by cybercriminals targeting organizations across various sectors [16].

Ransomware detection relies on two core analyses: static and dynamic. Static analysis examines code without execution, offering security but struggling with packed or obfuscated malware. Dynamic analysis executes code in a controlled environment, effectively capturing behavioural patterns but posing security risks. Often, a hybrid approach combines these methods with others to maximize accuracy and minimize associated risks.

This research paper is driven by three core objectives. Firstly, it aims to establish a robust system for the early identification of ransomware threats, prioritizing swift detection to minimize potential damage inflicted upon data and computer systems. Secondly, the paper endeavors to raise awareness among individuals and organizations regarding the substantial risks and dire consequences associated with

ransomware attacks, fostering a proactive and vigilant cybersecurity mindset. Lastly, a central focus of this research is the development and implementation of effective countermeasures and response strategies to mitigate the evolving menace of ransomware. By achieving these objectives, this work not only enhances our ability to combat ransomware but also contributes significantly to the broader realm of cybersecurity, fortifying our digital defences against this persistent and pernicious cyber threat.

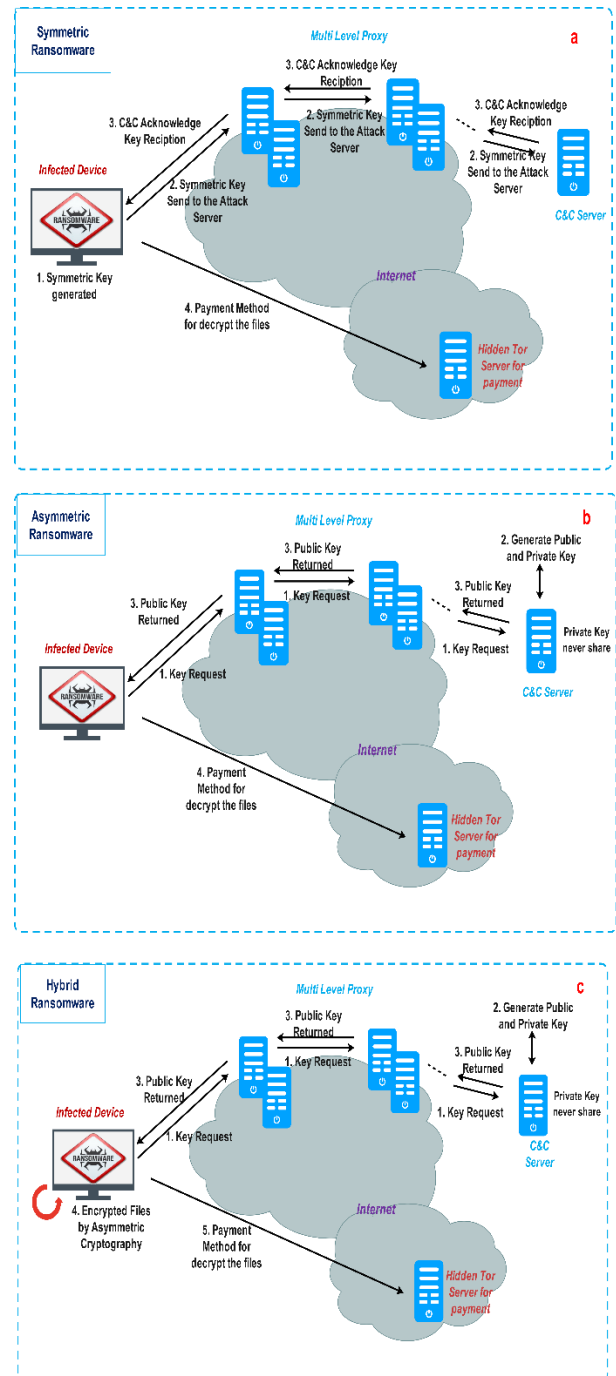


Fig. 1. Types of Ransomwares (a: Symmetric, b: Asymmetric, c: Hybrid).

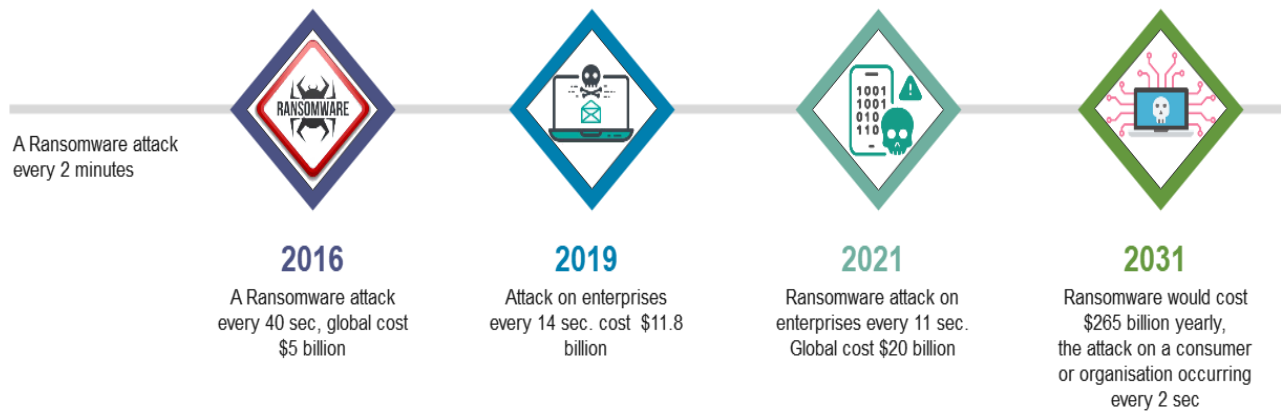


Fig. 2. DBIR and cybersecurity ventures report.

## II. RELATED WORK

Palisse et al. [17] recommended using Cryptographic approaches for proactively protecting against Ransomware, where ransomware constructs a dynamic block for calls to cryptographic APIs based on weaknesses and exploits vulnerabilities in cipher modes of operation. A mechanism termed PAYBREAK was suggested by Kolodenkerz et al. [18]. Because the system logs all of the random numbers it generates in a massive log file or database, users may use this information to exhaustively search for encryption keys. That method has proven to be quite effective. Kim et al. [19] used the same methods as the earlier researchers and developed a Deterministic Random Bit Generator (DRBG) to thwart ransomware; the DRBG is used to generate a seed, which is then combined with user DRBG data to produce a decryption key.

Poudyal et al. [20] suggested using reverse engineering to pull elements like assembly instructions and DLLs, which can subsequently be used with machine learning methods like K-fold Cross-Validation to identify ransomware. Ahmadian et al. [14] employed a system that relies on the Connection Monitor & Connection Breaker (CMCB) to identify and detect stealthy ransomware by monitoring and analysing network activity. Tseng et al. [21] and Cabaj et al. [22] conducted researches on the analysis and detection of ransomware attacks from different perspectives. They focused on analysing the HTTP and TCP protocols to identify ransomware attacks. Tseng utilised deep learning techniques for this purpose. On the other hand, Cabaj designed an approach based on Software-Defined Networking (SDN) to both detect and mitigate ransomware attacks. Cusack et al. [23] and Al Mashhadani et al. [8] proposed a framework for monitoring and analyzing data traffic during a ransomware attack. Their approach involved intercepting the communication between an infected machine, unaffected devices, and the command-and-control (C&C) server using a network protocol analyzer like Wireshark. To ensure the security of the monitoring system, they recommended implementing a firewall to shield it from potential threats. Shakir et al. [24] delved into the evolution of ransomware, tracing its development from phone antivirus and deceptive software to the emergence of crypto-ransomware. The study revealed two crucial factors driving the increase in ransomware attacks. Firstly, tracking victim payments to attackers proved to

be challenging due to the use of anonymous channels, making it challenging to trace and apprehend the perpetrators. Secondly, the practicality and effectiveness of employing cryptographic technologies played a significant role in the surge of ransomware attacks. These findings shed light on the key catalysts behind the proliferation and sophistication of ransomware as a malicious threat. The approach of Homayoun et al. [25] relied on the monitoring of activity records to detect and identify ransomware attacks by utilising machine learning methods to identify specific patterns. By analysing and examining activity logs, which capture system and user behaviour, it becomes possible to uncover anomalous patterns that indicate the presence of ransomware whereas Medhat et al. [26] developed a novel framework that relies on a static analysis approach. The framework utilises YARA rules, which are a set of predefined rules for pattern matching, to extract feature rules for each file. By applying these rules and conducting a classification process, the framework can identify files or processes that exhibit ransomware-like behaviour. The details and specific workings of this framework are represented in Fig. 3.

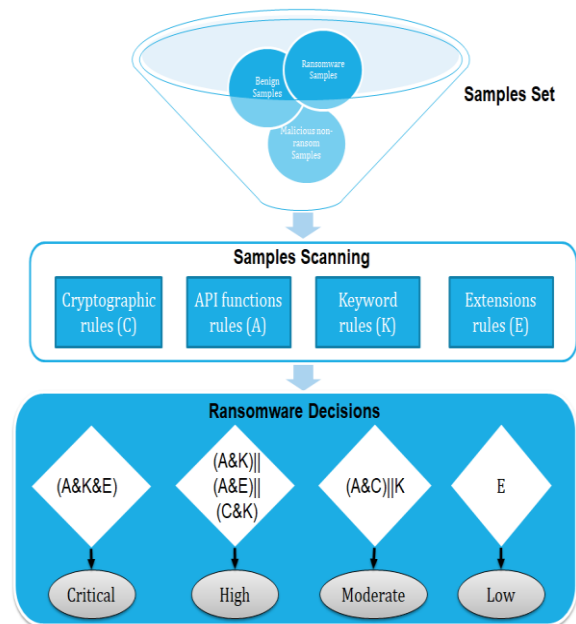


Fig. 3. Static-based framework for ransomware detection [25].

Detecting ransomware is a complex task due to its ever-changing nature, requiring researchers and security professionals to employ diverse strategies. One recommended approach involves monitoring the Master File Table (MFT) for any unusual activity or modifications, as ransomware often targets and encrypts files. Another approach involves leveraging machine learning models to search for specific language patterns associated with ransomware. By analyzing text data using these models, suspicious indicators can be identified. Additionally, examining data flow and network traffic can help detect ransomware [27]. Moore et al. [28] introduced a novel method for detecting ransomware utilising honeypot technology, as depicted in Fig. 4. This approach encompasses several stages to enhance detection capabilities. In the behaviour stage, two tools, namely DatAdvantage and HitmanPro, are utilised. DatAdvantage employs user behaviour analytics to identify abnormal activities, while HitmanPro focuses on detecting unusual system behaviour. The network stage involves monitoring data traffic across the network to detect any exchange of file keys, a common characteristic of ransomware operations. Lastly, in the server stage, changes are monitored using the file server resource manager, and a file screening function is employed to control access and block the writing of unauthorised files. This multi-stage approach aims to provide a comprehensive detection mechanism, combining behavioural analysis, network monitoring, and server-level control to enhance ransomware detection and prevention capabilities.

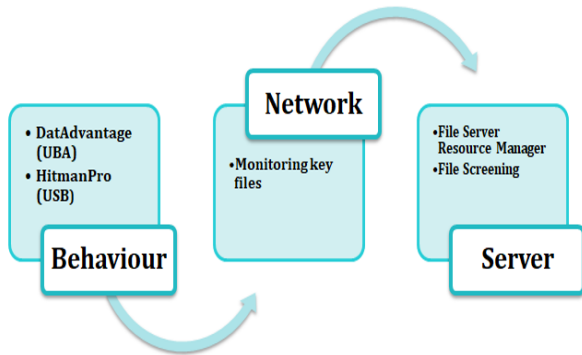


Fig. 4. Honeypot techniques to detect ransomware [28].

Ransomware can be classified into two main types: non-spreading ransomware and ransomware with worm-like characteristics. The primary objective in combating ransomware is to prevent the victim's device from being encrypted and to halt the malware's spread within and beyond the local network. Cabaj et al. [29] utilised SDN to detect non-spreading ransomware variants like CryptoWall and Locky. Their approach involved using the size of the first three HTTP Post packets in sequence as a detection mechanism. Through their implementation, they achieved a high true positive rate of 97-98%. In the case of ransomware that spreads like worms, such as WannaCry and ExPetr, researchers in [30] have suggested utilising two programmes for effective mitigation. The first programme detects and prevents WannaCry from encrypting the victim's device by monitoring and blocking communications between the ransomware and its Command-and-Control (C&C) servers through a dynamic IP blacklist.

The second programme tracks the ports used by WannaCry, enabling the prevention of encryption processes and the identification and containment of the ransomware's spreading behaviour.

### III. METHODOLOGY

The methodology (see Fig. 5) describes a technique for detecting ransomware using Cuckoo Sandbox [31], feature extraction from Portable Executable (PE) files [32], and YARA rules [33]. By analysing the characteristics and patterns of PE files, it is possible to identify potential ransomware threats and lessen their impact. There are three main stages to prepare data to be consumed by the algorithms we test: data collection, feature selection, and data processing. Data Collection is contingent upon collecting a diverse set of PE files, This investigation encompasses both benign components, which include application and system files that form the foundation of our digital activities, as well as malicious ransomware samples obtained from VirusShare, and utilising YARA rules to detect crypto signatures. Feature Selection is applied to analyse extracted characteristics and choose the most useful ones for ransomware detection. We employ Weight of Evidence (WoE) and Information Value (IV). They are used to determine a dataset's predictive potential and select which elements are most significant for a modelling task. Utilising data preprocessing techniques, such as normalisation, dimensionality reduction, data reshaping, and feature scaling is to improve the performance of subsequent machine learning algorithms.

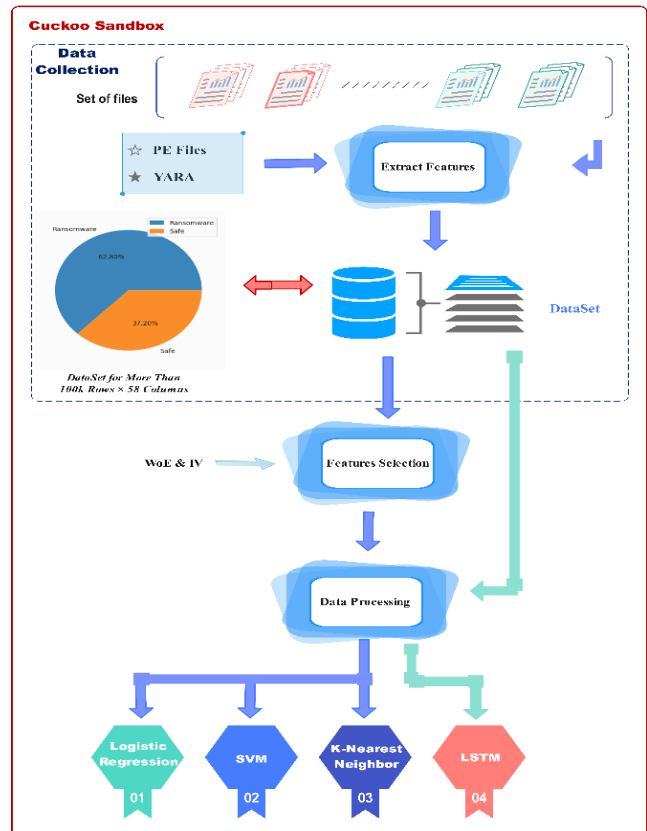


Fig. 5. Proposed methodology for detecting ransomware.

The output of the data processing stage is now suitable to be injected into the algorithms we test. We apply four machine learning algorithms: Logistic Regression, SVM, K-Nearest Neighbor, and LSTM.

Logistic Regression [34] is a type of generalized linear model that is used for classification tasks. The goal of logistic regression is to find the best model that describes the relationship between a set of input features and a binary output variable. The model is represented by a logistic function, which maps the input features to a value between 0 and 1, representing the probability that the output variable is in one of the two classes. In logistic regression, a linear combination of input features is transformed by the logistic function [35], which is also known as the sigmoid function. The coefficients of the linear combination are learned from the training data using a technique called maximum likelihood estimation. Once the model is trained, new data can be input into the model, and the output of the logistic function can be used to predict the probability that the new data belongs to one of the two classes. It is also less prone to overfitting compared to more complex models [36].

Support vector machines (SVMs) can solve classification and regression issues [37]. SVMs find the best boundary for categorising data. This border was chosen to maximise the margin, which is the distance between the boundary and each class's closest data points. After identifying this barrier, additional data may be categorised by its side [38]. SVMs are useful when data is not linearly separable, meaning classes cannot be separated by a straight line. SVMs use the "kernel trick" to shift data into a higher-dimensional space where it may be linearly segregated [39].

The K-Nearest Neighbors (K-NN) is a kind of supervised learning that classifies unknown files in line with previously established categories [40]. Among the many categorization algorithms, K-NN is a useful algorithm for grouping unknown objects into categories with the greatest number of shared attributes [41]. K-NN presupposes that close data points belong to the same class, which may not always be true. The data distribution, characteristics, and distance metric affect K-NN's efficacy [42].

Long-short term memory (LSTM), a standard and enhanced recurrent neural network, is capable of analysing and anticipating time series problems. A memory cell composed of an input gate, a forget gate, and an output gate regulates the transmission of information in the LSTM model. The problem of expanding gradients and vanishing is resolved by LSTM's unique structure [43].

#### IV. EVALUATION AND RESULTS

In order to evaluate the performance of classification models, various evaluation measures are commonly utilized. These measures include accuracy, precision, recall, and the F1 Score. The confusion matrix, represented as Fig. 6, is used to assess the classifier's primary performance indicators.

Accuracy, as expressed by Eq. (1) represents the percentage of correctly classified instances in relation to the entire dataset. Precision, or Detection Rate, as given by Eq. (2), is commonly used when dealing with imbalanced datasets. It measures the proportion of correctly classified instances compared to the total instances that are correctly and incorrectly classified. Recall, described in Eq. (3), is the percentage of accurately classified instances in relation to the total number of actual positive instances. The False Alarm Rate, mentioned in Eq. (4), indicates the frequency with which attacks are misclassified or falsely identified. Finally, the F1 Score or F-measure, described by Eq. (5), provides an overall assessment of the accuracy of a test by combining precision and recall into a single metric. This provides insights into the classifier's primary performance indicators and facilitates further analysis and comparison of different models and techniques.

		Predicted	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN

Fig. 6. Confusion matrix.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$False\_Alarm\_Rate = \frac{FP}{FP+TN} \quad (4)$$

$$F1\_Score = 2 * \frac{Precision*Recall}{Precision+Recall} \quad (5)$$

In feature selection, we have used WOE and IV as criteria for ranking and selecting variables. Variables with high IV values are considered more predictive and are more likely to be included in the final set of features, as shown in Fig. 7. By focusing on variables with strong predictive power, it can reduce dimensionality and improve the performance and interpretability of the used models.

Table I and Fig. 8 present a comprehensive overview of the performance results of different models utilized for ransomware detection, including LSTM, SVM, LR, and KNN. These findings offer valuable insights into the capabilities and effectiveness of each model in addressing the task at hand. Since the dataset consists of sequences of bits, LSTM's ability to remember and learn from previous information makes it a valuable choice. By utilizing its memory cells and gating mechanisms, the LSTM model can effectively process and interpret the sequential nature of the data, making it well-suited for the task at hand.

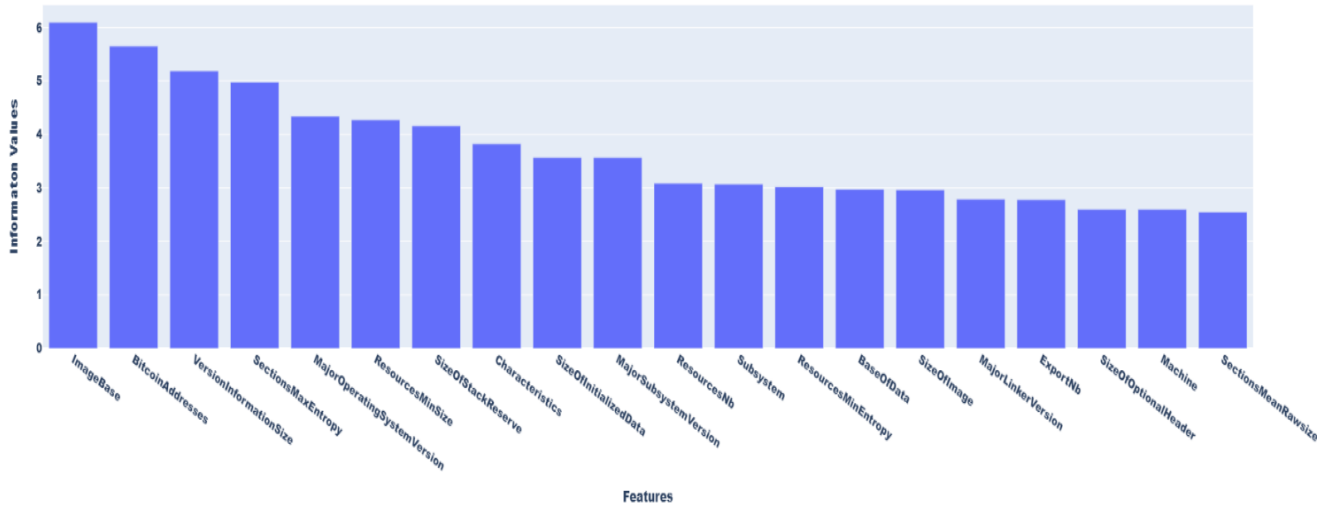


Fig. 7. Importance of features based on information value.

TABLE I. EXPERIMENTAL RESULTS

	LSTM	SVM	LR	KNN
TN	28689	28651	28342	28380
FN	235	283	592	494
FP	146	378	640	281
TP	12345	12103	11841	12260
Total	41415	41415	41415	41415
Accuracy	<b>0.9908</b>	<b>0.98404</b>	<b>0.970252</b>	<b>0.981287</b>

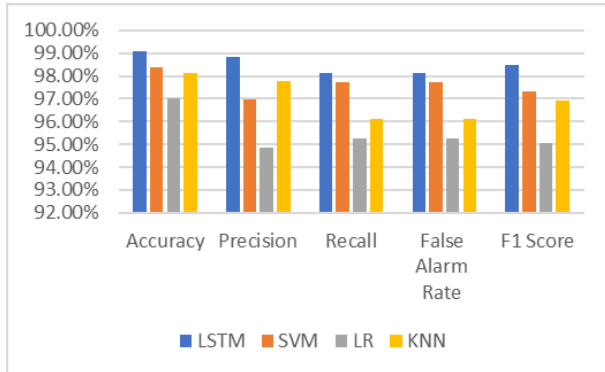


Fig. 8. Comprised the results of applied algorithms.

Support Vector Machine (SVM) and logistic regression face challenges when dealing with massive datasets due to their computational complexity, leading to longer training times as the dataset size increases. However, for smaller datasets and lower-dimensional feature spaces, K-Nearest Neighbors (K-NN) remains a valuable and effective classification technique, boasting an impressive accuracy of 98%. K-NN's simplicity and ability to capture patterns make it a practical choice for such scenarios. The LSTM model exhibited the highest accuracy, achieving an impressive score of 0.9908. This indicates that the LSTM model is highly effective in detecting ransomware attacks, showcasing its ability to capture intricate patterns and temporal dependencies within the data. With such a high accuracy, the LSTM model can be considered a robust

choice for ransomware detection. During the training process, the LSTM model continuously updates its internal parameters to minimize the loss, resulting in better predictions and higher accuracy over time. As the model converges, the accuracy typically improves, and the loss decreases as shown in Fig. 9.

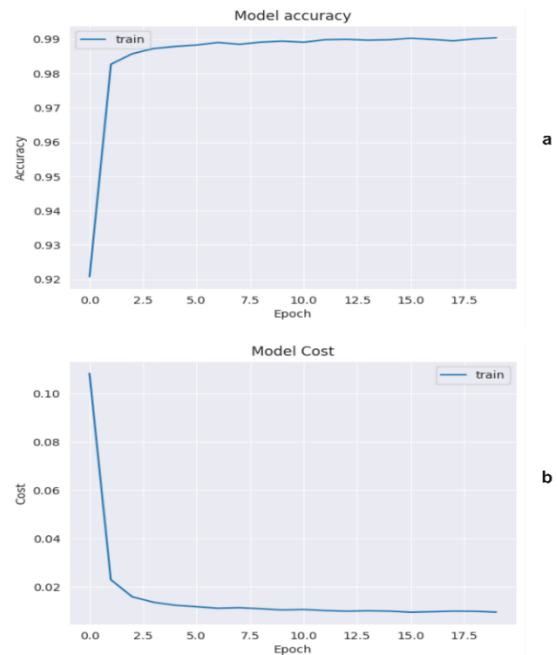


Fig. 9. Visualized results of LSTM (a. Accuracy & b. Loss mode).

## V. CONCLUSION AND FUTURE WORKS

In conclusion, the studies on ransomware detection have generally yielded positive results. Researchers have identified distinct static features in ransomware samples, such as cryptographic signatures, API methods, and file extensions, that can aid in their identification. Different approaches, including the use of surveillance and honeypots, have been explored to track down and analyze ransomware. Machine learning and classification techniques have proven to be valuable in enhancing both static and dynamic malware analysis.

Additionally, the application of deep learning, such as the LSTM model, has shown remarkable accuracy, reaching up to 99%, in detecting malware and ransomware. These findings highlight the potential of advanced techniques for improving cybersecurity measures against ransomware threats. Continued research and development in this field can further strengthen the detection and mitigation of ransomware attacks. In future work, we propose extending the dataset used for ransomware detection to include a more comprehensive set of features derived from both dynamic and static analysis. Moreover, we intend to explore the use of a hybrid algorithm combining CNN-LSTM models. This fusion of techniques has the potential to improve the accuracy and robustness of ransomware detection, paving the way for more effective defense mechanisms against evolving ransomware threats.

#### REFERENCES

- [1] M. Robinson, K. Jones, and H. Janicke, "Cyber warfare: Issues and challenges," *Computers & Security*, vol. 49, pp. 70–94, Mar. 2015, doi: <https://doi.org/10.1016/j.cose.2014.11.007>.
- [2] T. J. Holt and A. M. Bossler, *Cybercrime in progress : theory and prevention of technology-enabled offenses*. London ; New York: Routledge, 2016.
- [3] E. Filiol, M. Helenius, and S. Zanero, "Open Problems in Computer Virology," *Journal in Computer Virology*, vol. 1, no. 3–4, pp. 55–66, Feb. 2006, doi: <https://doi.org/10.1007/s11416-005-0008-3>
- [4] M. Sikorski and A. Honig, *Practical malware analysis : the hands-on guide to dissecting malicious software*. San Francisco No Starch Press, 2012.
- [5] T. Mane, Prachi Nimase, Prahalad Parihar, and Pragati Chandankhede, "Review of Malware Detection Using Deep Learning," *Oct. 2021*, doi: [https://doi.org/10.1007/978-981-16-5301-8\\_19](https://doi.org/10.1007/978-981-16-5301-8_19).
- [6] J. Rutkowska, "Introducing Stealth Malware Taxonomy," 2006.
- [7] A. Razgallah, R. Khoury, S. Hallé, and K. Khanmohammadi, "A survey of malware detection in Android apps: Recommendations and perspectives for future research," *Computer Science Review*, vol. 39, p. 100358, Feb. 2021, doi: <https://doi.org/10.1016/j.cosrev.2020.100358>.
- [8] A. O. Almashhadani, M. Kaiiali, S. Sezer, and P. O’Kane, "A Multi-Classier Network-Based Crypto Ransomware Detection System: A Case Study of Locky Ransomware," *IEEE Access*, vol. 7, pp. 47053–47067, 2019, doi: <https://doi.org/10.1109/access.2019.2907485>.
- [9] Monika, P. Zavorsky, and D. Lindskog, "Experimental Analysis of Ransomware on Windows and Android Platforms: Evolution and Characterization," *Procedia Computer Science*, vol. 94, pp. 465–472, 2016, doi: <https://doi.org/10.1016/j.procs.2016.08.072>.
- [10] K. Lee, K. Yim, and J. T. Seo, "Ransomware prevention technique using key backup," *Concurrency and Computation: Practice and Experience*, vol. 30, no. 3, p. e4337, Oct. 2017, doi: <https://doi.org/10.1002/cpe.4337>.
- [11] P. O’Kane, S. Sezer, and D. Carlin, "Evolution of ransomware," *IET Networks*, vol. 7, no. 5, pp. 321–327, 2018.
- [12] H. Orman, "Evil Offspring - Ransomware and Crypto Technology," *IEEE Internet Computing*, vol. 20, no. 5, pp. 89–94, Sep. 2016, doi: <https://doi.org/10.1109/mic.2016.90>.
- [13] N. Aldaraani and Z. Begum, "Understanding the impact of Ransomware: A Survey on its Evolution, Mitigation and Prevention Techniques," *IEEE Xplore*, pp. 1–5, Apr. 2018, doi: <https://doi.org/10.1109/NCG.2018.8593029>.
- [14] M. M. Ahmadian, H. R. Shahriari, and S. M. Ghaffarian, "Connection-monitor & connection-breaker: A novel approach for prevention and detection of high survivable ransowares," 2015 12th International Iranian Society of Cryptology Conference on Information Security and Cryptology (ISCISC), Sep. 2015, doi: <https://doi.org/10.1109/iscisc.2015.7387902>.
- [15] A. Liska and T. Gallo, *Ransomware : defending against digital extortion*. Sebastopol (Calif.): O’reilly Media. Copyright, 2016.
- [16] A. Fagioli, "Zero-day recovery: the key to mitigating the ransomware threat," *Computer Fraud & Security*, vol. 2019, no. 1, pp. 6–9, Jan. 2019, doi: [https://doi.org/10.1016/s1361-3723\(19\)30006-5](https://doi.org/10.1016/s1361-3723(19)30006-5).
- [17] A. Palisse, H. Le Boudier, J.-L. Lanet, C. Le Guernic, and A. Legay, "Ransomware and the Legacy Crypto API," *Lecture Notes in Computer Science*, vol. 10158, pp. 11–28, 2017, doi: [https://doi.org/10.1007/978-3-319-54876-0\\_2](https://doi.org/10.1007/978-3-319-54876-0_2).
- [18] E. Kolodenker, W. Koch, G. Stringhini, and M. Egele, "PayBreak : Defense Against Cryptographic Ransomware," *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security*, Apr. 2017, doi: <https://doi.org/10.1145/3052973.3053035>.
- [19] H. Kim, D. Yoo, J. -S. Kang and Y. Yeom, "Dynamic ransomware protection using deterministic random bit generator," 2017 IEEE Conference on Application, Information and Network Security (AINS), Miri, Malaysia, 2017, pp. 64–68, doi: [10.1109/AINS.2017.8270426](https://doi.org/10.1109/AINS.2017.8270426).
- [20] S. Poudyal, K. P. Subedi, and D. Dasgupta, "A Framework for Analyzing Ransomware using Machine Learning," *IEEE Xplore*, Nov. 01, 2018.
- [21] A. Tseng, Y. Chen, Y. Kao, and T. Lin, "Deep Learning for Ransomware Detection," *IEICE Technical Report; IEICE Tech. Rep.*, vol. 116, no. 282, pp. 87–92, Oct. 2016,
- [22] K. Cabaj and W. Mazurczyk, "Using Software-Defined Networking for Ransomware Mitigation: The Case of CryptoWall," *IEEE Network*, vol. 30, no. 6, pp. 14–20, Nov. 2016, doi: <https://doi.org/10.1109/mnet.2016.1600110nm>.
- [23] G. Cusack, O. Michel, and E. Keller, "Machine Learning-Based Detection of Ransomware Using SDN," *Proceedings of the 2018 ACM International Workshop on Security in Software Defined Networks & Network Function Virtualization*, pp. 1–6, Mar. 2018, doi: <https://doi.org/10.1145/3180465.3180467>.
- [24] H. A. Shakir and A. N. Jaber, "A Short Review for Ransomware: Pros and Cons," *Advances on P2P, Parallel, Grid, Cloud and Internet Computing*, pp. 401–411, Nov. 2017, doi: [https://doi.org/10.1007/978-3-319-69835-9\\_38](https://doi.org/10.1007/978-3-319-69835-9_38).
- [25] S. Homayoun, A. Dehghantanha, M. Ahmadzadeh, S. Hashemi, and R. Khayami, "Know Abnormal, Find Evil: Frequent Pattern Mining for Ransomware Threat Hunting and Intelligence," *IEEE Transactions on Emerging Topics in Computing*, vol. 8, no. 2, pp. 341–351, Apr. 2020, doi: <https://doi.org/10.1109/tetc.2017.2756908>.
- [26] M. Medhat, S. Gaber, and N. Abdelbaki, "A New Static-Based Framework for Ransomware Detection," 2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress(DASC/PiCom/DataCom/CyberSciTech), Aug. 2018, doi: <https://doi.org/10.1109/dasc/picom/datacom/cyberscitech.2018.00124>.
- [27] N. Andronio, S. Zanero, and F. Maggi, "HelDroid: Dissecting and Detecting Mobile Ransomware," *Research in Attacks, Intrusions, and Defenses*, pp. 382–404, 2015, doi: [https://doi.org/10.1007/978-3-319-26362-5\\_18](https://doi.org/10.1007/978-3-319-26362-5_18).
- [28] C. Moore, "Detecting Ransomware with Honeypot Techniques," 2016 Cybersecurity and Cyberforensics Conference (CCC), Aug. 2016, doi: <https://doi.org/10.1109/cc.2016.14>.
- [29] K. Cabaj, M. Gregorczyk, and W. Mazurczyk, "Software-defined networking-based crypto ransomware detection using HTTP traffic characteristics," *Computers & Electrical Engineering*, vol. 66, pp. 353–368, Feb. 2018, doi: <https://doi.org/10.1016/j.compeleceng.2017.10.012>.
- [30] M. Akbanov, V. G. Vassilakis, and M. D. Logothetis, "Ransomware detection and mitigation using software-defined networking: The case of WannaCry," *Computers & Electrical Engineering*, vol. 76, pp. 111–121, Jun. 2019, doi: <https://doi.org/10.1016/j.compeleceng.2019.03.012>.
- [31] L. Wang, B. Wang, J. Liu, Q. Miao, and J. Zhang, "Cuckoo-based Malware Dynamic Analysis," *International Journal of Performability Engineering*, 2019, doi: <https://doi.org/10.23940/ijpe.19.03.p6.772781>.
- [32] Y. Zhang, X. Chang, Y. Lin, J. Mistic, and V. B. Mistic, "Exploring Function Call Graph Vectorization and File Statistical Features in Malicious PE File Classification," *IEEE Access*, vol. 8, pp. 44652–44660, 2020, doi: <https://doi.org/10.1109/access.2020.2978335>.



- [33] N. Naik et al., "Embedded YARA rules: strengthening YARA rules utilising fuzzy hashing and fuzzy rules for malware analysis," *Complex & Intelligent Systems*, vol. 7, no. 2, pp. 687–702, Nov. 2020, doi: <https://doi.org/10.1007/s40747-020-00233-5>.
- [34] J. M. Hilbe, *Logistic Regression Models*. Chapman and Hall/CRC, 2009. doi: <https://doi.org/10.1201/9781420075779>.
- [35] Z. Akram, M. Majid, and S. Habib, "A Systematic Literature Review: Usage of Logistic Regression for Malware Detection," *IEEE Xplore*, Nov. 01, 2021. <https://ieeexplore.ieee.org/document/9693035> (accessed Jan. 12, 2023).
- [36] Ö. A. Aslan and R. Samet, "A Comprehensive Review on Malware Detection Approaches," *IEEE Access*, vol. 8, pp. 6249–6271, Jan. 2020, doi: <https://doi.org/10.1109/ACCESS.2019.2963724>.
- [37] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *Neurocomputing*, vol. 408, pp. 189–215, Sep. 2020, doi: <https://doi.org/10.1016/j.neucom.2019.10.118>.
- [38] M. Wadkar, F. Di Troia, and M. Stamp, "Detecting malware evolution using support vector machines," *Expert Systems with Applications*, vol. 143, p. 113022, Apr. 2020, doi: <https://doi.org/10.1016/j.eswa.2019.113022>.
- [39] P. O’Kane, S. Sezer, K. McLaughlin, and E. G. Im, "SVM Training Phase Reduction Using Dataset Feature Filtering for Malware Detection," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 3, pp. 500–509, Mar. 2013, doi: <https://doi.org/10.1109/tifs.2013.2242890>.
- [40] F. A. Narudin, A. Feizollah, N. B. Anuar, and A. Gani, "Evaluation of machine learning classifiers for mobile malware detection," *Soft Computing*, vol. 20, no. 1, pp. 343–357, Nov. 2014, doi: <https://doi.org/10.1007/s00500-014-1511-6>.
- [41] D. Stiawan, S. M. Daely, A. Heryanto, N. Afifah, M. Y. Idris, and R. Budiarto, "Ransomware Detection Based On Opcode Behavior Using K-Nearest Neighbors Algorithm," *Information Technology and Control*, vol. 50, no. 3, pp. 495–506, Sep. 2021, doi: <https://doi.org/10.5755/j01.itc.50.3.25816>.
- [42] H. A. Abu Alfeilat et al., "Effects of Distance Measure Choice on K-Nearest Neighbor Classifier Performance: A Review," *Big Data*, vol. 7, no. 4, pp. 221–248, Dec. 2019, doi: <https://doi.org/10.1089/big.2018.0175>.
- [43] R. Lu, "Malware Detection with LSTM using Opcode Language," arXiv:1906.04593 [cs], Jun. 2019

# K-Means Extensions for Clustering Categorical Data on Concept Lattice

Mohammed Alwersh\*, László Kovács

Department of Information Technology, University of Miskolc, Miskolc, Hungary

**Abstract**—Formal Concept Analysis (FCA) is a key tool in knowledge discovery, representing data relationships through concept lattices. However, the complexity of these lattices often hinders interpretation, prompting the need for innovative solutions. In this context, the study proposes clustering formal concepts within a concept lattice, ultimately aiming to minimize lattice size. To address this, the study introduces two novel extensions of the k-means algorithm to handle categorical data efficiently, a crucial aspect of the FCA framework. These extensions, namely K-means Dijkstra on Lattice (KDL) and K-means Vector on Lattice (KVL), are designed to minimize the concept lattice size. However, the current study focuses on introducing and refining these new methods, laying the groundwork for our future goal of lattice size reduction. The KDL utilizes FCA to build a graph of formal concepts and their relationships, applying a modified Dijkstra algorithm for distance measurement, thus replacing the Euclidean distance in traditional k-means. The defined centroids are formal concepts with minimal intra-cluster distances, enabling effective categorical data clustering. In contrast, the KVL extension transforms formal concepts into numerical vectors to leverage the scalability offered by traditional k-means, potentially at the cost of clustering quality due to oversight of the data's inherent hierarchy. After rigorous testing, KDL and KVL proved robust in managing categorical data. The introduction and demonstration of these novel techniques lay the groundwork for future research, marking a significant stride toward addressing current challenges in categorical data clustering within the FCA framework.

**Keywords**—Clustering algorithms; categorical data; k-means; cluster analysis; formal concept analysis; concept lattice

## I. INTRODUCTION

The technique of data clustering involves an unsupervised classification method that aims to group a set of unlabeled objects into meaningful clusters based on their similarities and differences. This process requires objects within the same cluster to exhibit high similarity, while objects in different clusters should have significant differences. Similarities and dissimilarities between objects are evaluated by considering attribute values that describe the objects, often using distance measures. Typically, objects can be represented as vectors in a multidimensional space, with each dimension representing a feature. When numerical features describe objects, geometric distance measures such as Euclidean or Manhattan distance can be used to define their similarity. However, these distance measures are unsuitable for categorical data, including values like gender or location. In recent years, there has been increasing interest in clustering categorical data [1-3]. For categorical data, a comparison measure, rather than a distance

measure, is commonly used [4]. However, this metric does not differentiate between different attribute values since it only measures equality between pairs of values [5].

The widely acclaimed k-means algorithm [6] excels in simplicity and efficiency, particularly for large numerical datasets. However, it is restricted by its inability to handle categorical data directly. To overcome this limitation, adaptations such as the k-modes [3], k-representative [7], and k-centers algorithms [8] have been introduced. The k-modes algorithm uses simple matching similarity measures and substitutes "means" with "modes" for cluster centers. This modification, however, can result in multiple modes within a cluster, which can influence the algorithm's performance. To navigate this, the k-representative algorithm proposed in [9], presents "cluster centers" uniquely suited to categorical data. In this approach, the defining attribute of a cluster's representative stems from the diversity of categorical values within the cluster itself [9]. Although these adaptations of the k-means algorithm share a similar clustering process, they define "cluster center" and "similarity measure" differently for categorical data, which could potentially result in information loss when categorical data are directly transformed into vector space.

This paper introduces two novel extensions of the k-means algorithm for clustering categorical data: K-means Dijkstra on Lattice (KDL) and K-means Vector on Lattice (KVL). Existing methods such as k-means and its variants are restricted by their inability to adequately handle categorical data, often requiring data transformation techniques that can result in information loss. Additionally, they struggle with representing complex, hierarchical relationships inherent in the data. Our proposed methods, KDL and KVL, aim to overcome these limitations by focusing on learning and integrating the representation of categorical values based on their inherent graph structure. This approach not only preserves the richness of the categorical data but also effectively captures the potential similarity and hierarchical relationships between the values. KDL utilizes Formal Concept Analysis to construct a graph, with nodes symbolizing formal concepts and edges representing hierarchical relationships [10]. A customized Dijkstra algorithm identifies the shortest path between formal concepts, replacing the Euclidean distance from traditional k-means. The centroids are those formal concepts with the least sum of intra-cluster distances. This method accurately identifies data patterns and relationships, providing insights often missed by conventional vector representations. KVL, the second extension, transforms formal concepts into numerical concept description vectors and applies traditional k-means. It evaluates concept similarity, groups related concepts, and positions each

cluster's center as the mean of its concept description vectors. This approach excels in scalability by converting categorical data into numerical formats, enabling efficient analysis of larger datasets. Both KDL and KVL significantly advance our understanding of complex datasets by providing a unique perspective on clustering categorical data. These novel extensions of the k-means algorithm pave the way for further research in this field. The primary contributions of this study are the introduction of KDL and KVL as innovative extensions of the k-means algorithm, the comparative evaluation of these methods against existing algorithms, and the laying of groundwork for future advancements in clustering categorical data within the Formal Concept Analysis frameworks.

The remainder of this paper is structured as follows: Section II provides an overview of Formal Concept Analysis (FCA), elucidating key terminologies and notions. Section III delves into Dijkstra's Algorithm and its variations employed for addressing shortest-path problems. Section IV then explores the k-means algorithm and its extended versions tailored for categorical data. This segues into Section V, where the proposed clustering methodologies are thoroughly discussed. The experimental results and their implications are exhibited in Section VI. The paper concludes in Section VII, summarizing the study's main points and potential future directions.

## II. FCA: KEY TERMINOLOGY AND NOTIONS

Formal Concept Analysis (FCA) emerged as a distinctive mathematical field in the early 1980s. Central to its application are particular diagrams known as line or Hasse diagrams, which are utilized to depict information via concept lattices [10]. To enhance the comprehension of the study, the study provide a succinct overview that delineates the critical concepts and definitions within FCA, supplemented by an easily digestible example. The terms and foundations of FCA discussed in this paper are grounded in the work of [11].

**Definition 1. Formal Context:** A formal context is characterized as a tripartite structure  $(G, M, I)$ , where  $G$  represents a collection of objects,  $M$  stands for a set of attributes, and  $I \subseteq G \times M$ , embodies the incidence relationship between  $G$  and  $M$ . For each object  $g \in G$  and attribute  $m \in M$ , a binary relation  $gIm ((g, m) \in I)$  signifies that object  $g$  possesses attribute  $m$ . Typically, a cross table illustrates a formal context, where rows depict the object names and columns exhibit the attribute names. The presence or absence of a cross demonstrates the existence or non-existence of an incidence relationship between  $G$  and  $M$ , respectively. This context is also referred to as a binary context, as demonstrated in Table I.

Expanding on Table I shows a compact formal context in which the object set  $G$  includes  $\{o_1, o_2, o_3, o_4, o_5\}$ , and the attribute set  $M$  comprises  $\{a_1, a_2, a_3, a_4, a_5, a_6\}$ . The cross (x) at the intersection of the object  $g$  row and the attribute  $m$  column indicates that the object  $g$  possesses the attribute  $m$ , while its absence signifies a lack of relationship between  $g$  and  $m$ . For instance, within this formal context, object  $o_1$  is associated with the attributes  $\{a_2, a_6\}$ .

**Definition 2. Derivation Operators:** In a formal context  $(G, M, I)$ , derivation operators  $\uparrow = 2^G \rightarrow 2^M$ ,  $\downarrow = 2^M \rightarrow 2^G$ , are established for every subset of objects  $A$  within  $G$  and a subset of attributes  $B$  within  $M$ . They are precisely defined as  $A^\uparrow = \{m \in M \mid \forall g \in A: (g, m) \in I\}$ ,  $B^\downarrow = \{g \in G \mid \forall m \in B: (g, m) \in I\}$ . The upward operator  $A^\uparrow$  represents the collective attributes shared by all objects in  $A$ , and the downward operator  $B^\downarrow$  comprises all objects that possess all attributes in  $B$ .

These derivation operators  $A^\uparrow$  and  $B^\downarrow$  are also referred to as  $A'$  and  $B'$ . For example, within the context provided in Table I, it can be easily discerned that:

$$\{o_1\}^\uparrow = \{a_2, a_6\}$$

$$\{o_1, o_3\}^\uparrow = \{a_6\}$$

$$\{a_1\}^\downarrow = \{o_3\}$$

$$\{a_2, a_3\}^\downarrow = \{o_5\}$$

**Definition 3. Formal Concepts:** In a provided context denoted as  $k=(G, M, I)$ , a formal concept is recognized as a pair  $(A, B)$ , where  $A$ , a subset of  $G$ , is referred to as the 'extent' part of the formal concept  $(A, B)$ , while  $B$ , a subset of  $M$ , is known as the 'intent' part of the formal concept  $(A, B)$ , provided that  $A' = B, B' = A$ . For instance, considering the formal context illustrated in Table I, the pair  $(\{o_1, o_3\}, \{a_6\})$  emerges as a formal concept, where  $\{o_1, o_3\}$  represents the extent part and  $\{a_6\}$  embodies the intent part.

**Definition 4. Concept Lattices:** Formal concepts can be arranged based on the subconcept-super concept relation  $\leq$ , expressed as follows:  $(A_1, B_1) \leq (A_2, B_2) \Leftrightarrow A_1 \subseteq A_2$  (or equivalently  $B_1 \subseteq B_2$ ), where  $(A_1, B_1)$  is a subconcept (more specific) and  $(A_2, B_2)$  is a super concept (more general). Within a formal context  $K$ , the assembly of all formal concepts, in conjunction with the partial order  $\leq$ , generally defines a complete lattice, referred to as a concept lattice [10]. The notation  $\mathcal{B}(G, M, I)$  signifies the concept lattice derived from a formal context  $(G, M, I)$ .

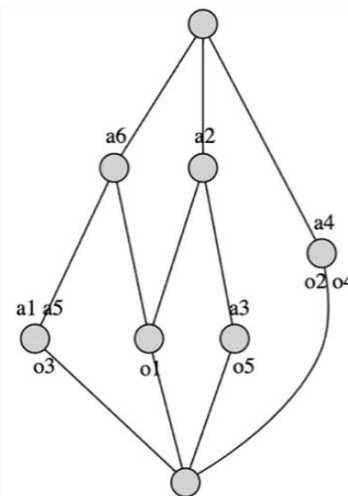


Fig. 1. Concept lattice corresponding to the formal context from Table I.

By the initial part of the core theorem on concept lattices [10], a concept lattice  $\mathcal{B}(G, M, I)$  is identified as a complete lattice where the infimum and supremum are present for any arbitrary set. This is represented as  $(A_1, B_1) \wedge (A_2, B_2) = (A_1 \cap B_2, (B_1 \cup B_2)'')$  and  $(A_1, B_1) \vee (A_2, B_2) = ((A_1 \cup A_2)'', B_1 \cap B_2)$ .

TABLE I. FORMAL CONTEXT

Objects/ Attributes	a <sub>1</sub>	a <sub>2</sub>	a <sub>3</sub>	a <sub>4</sub>	a <sub>5</sub>	a <sub>6</sub>
o <sub>1</sub>		x				x
o <sub>2</sub>				x		
o <sub>3</sub>	x				x	x
o <sub>4</sub>				x		
o <sub>5</sub>		x	x			

The concept lattice derived from the formal context showcased in Table I is portrayed in Fig. 1 through a line diagram. The lattice is constituted by formal concepts, which are generated depending on the formal context and their relationship defined by the subconcept-superconcept paradigm [11]. Every node in the line diagram symbolizes a formal concept. These formal concepts in the diagram can be categorized into object concepts, described as  $(\{g\}'', \{g\}')$ , or attribute concepts, defined as  $(\{m\}', \{m\}'')$ , indicating the concepts tied to objects or attributes, respectively. An object concept is represented as  $\gamma(g)$ , while an attribute concept is signified as  $\mu(m)$ . When labeling  $\mathcal{B}(G, M, I)$ , The convention adhered to in this study specifies that each object  $g$  gets its label  $g$  for the corresponding formal concept  $\gamma(g)$ , and each attribute  $m$  is labeled as  $m$  for the formal concept  $\mu(m)$ . Certain concept nodes carry objects below them, while others have attributes above them. More often than not, object labeling is placed underneath a node, while attribute labeling is positioned above it in a line diagram. It's worth highlighting that not every concept within a particular context is necessarily an object or an attribute concept. Any concept has the potential to be an object concept, an attribute concept, a blend of the two, or neither [12,13].

### III. DIJKSTRA'S ALGORITHM AND ITS VARIATIONS FOR SHORTEST PATH PROBLEMS

Dijkstra's algorithm was introduced in 1959 by Dutch computer scientist Edsger W. Dijkstra; Dijkstra's algorithm serves as a leading method for finding optimal paths. This algorithm, notable for its efficiency and effectiveness, has found broad application across various fields. One crucial application is in routing through the Internet's vast interconnected web, determining the shortest paths between network nodes. The algorithm operates on a directed weighted graph denoted as  $G = \{V, E\}$ , where  $V$  is the set of vertices and  $E$  the set of edges. Each edge  $e$  has an associated weight, Edge-Cost ( $e$ ), representing the traversal cost. It's essential for the algorithm's efficiency that all weights are non-negative. By applying Dijkstra's algorithm, one can efficiently ascertain the shortest path from a source vertex to every other vertex in the graph. Such information finds applications in diverse fields, including transportation, logistics, and network communication [14].

The algorithm proceeds by assigning temporary and visited states to vertices and updating their distances from the source. It marks vertices as visited, updating the temporary vertices' distances as it progresses. After visiting all vertices, the algorithm terminates. However, it has limitations: it involves a blind search, leading to inefficient resource use and longer operations. Moreover, it cannot handle negative edges, leading to potential inaccuracies in shortest path calculations [15]. The efficiency of Dijkstra's algorithm can be expressed using Big-O notation. The complexity depends on the number of vertices ( $|V| = n$ ) and the updates for priority queues ( $|E|$ ). Different data structures can be used for the priority queue, resulting in different complexities:

- If Fibonacci heap is used, the complexity is  $O(|E| + |V| \log |V|)$ , with DeleteMin operations taking  $O(\log |V|)$ .
- If a standard binary heap is used, the complexity is  $O(|E| \log |E|)$ , with  $|E|$  updates for the standard heap.
- If a priority queue is used, the complexity is  $O(|E| + |V|^2)$ , where the  $|V|^2$  term arises from  $|V|$  scans of the unordered set New Frontier to find the vertex with the least sDist value.

There are numerous variants of Dijkstra's algorithm, each addressing specific needs and modifications. Among the most well-known is the Bellman-Ford algorithm [16], which caters to negative-weighted graphs. The Floyd-Warshall algorithm [17], employs dynamic programming to find shortest paths, accommodating both positive and negative edge weights. The Johnson algorithm uses the Bellman-Ford algorithm to re-weight the graph, eliminating negative weights, then applies Dijkstra's algorithm, effectively reducing execution time for sparse graphs [18]. Lastly, the A\* algorithm extends Dijkstra's by combining breadth-first search and heuristic methods, potentially increasing speed but failing to ensure absolute accuracy [14]. Depending on the problem's requirements and characteristics, one can choose the most suitable variant to obtain optimal results. Dijkstra's algorithm is utilized to measure distance in the particular use case of clustering categorical data within a concept lattice. This approach allows us to interpret the intricate structures of the lattice effectively, paving the way for efficient clustering and insight generation.

### IV. K-MEANS ALGORITHM AND ITS EXTENSIONS FOR CATEGORICAL DATA

The k-means algorithm [6], is a widely used partitional or non-hierarchical clustering method. Given a set  $D$  of  $N$  numerical data objects, a natural number  $k$  (where  $k$  is less than or equal to  $N$ ), and a distance measure between objects, the algorithm aims to find a partition of  $D$  into  $k$  non-empty and disjoint clusters. The objective is to minimize the sum of squared distances between each data object and its assigned cluster center.

Mathematically, the k-means algorithm can be formulated as an optimization problem. Let  $U = [u_{i,j}]$  be the partition matrix, where  $u_{i,j}$  is a binary indicator variable representing whether object  $X_i$  belongs to cluster  $S_j$ . Let  $Z = \{Z_1, Z_2, \dots, Z_k\}$  be the set of cluster centers. The squared

Euclidean distance  $dis(\cdot, \cdot)$  between two objects  $X_i$  and  $Z_j$  is used as the distance measure [19].

The objective function to be minimized is given by  $P$  problem, as specified in Eq. (1):

$$P(U, Z) = \sum_{j=1}^k \sum_{i=1}^N u_{i,j} dis(X_i, Z_j) \quad (1)$$

The objective function presented in Eq. (1) is to be minimized, and this operation is constrained by specific conditions as outlined in Eq. (2):

$$\begin{aligned} \sum_{j=1}^k u_{i,j} &= 1, \quad 1 \leq i \leq N. \\ u_{i,j} &\in \{0,1\}, \quad 1 \leq i \leq N, 1 \leq j \leq k \end{aligned} \quad (2)$$

where  $U = [u_{i,j}]_{N \times k}$  represent a partition matrix, where each element  $u_{i,j}$  equals 1 if object  $X_i$  belongs to cluster  $S_j$  and 0 otherwise.  $Z = \{Z_1, Z_2, \dots, Z_k\}$  denotes the set of cluster centers. The function  $dis(\cdot, \cdot)$  calculates the squared Euclidean distance between two objects.

The k-means algorithm follows a four-step process until the objective function  $P(U, Z)$  converges to a local minimum:

- Initialize cluster centers as  $Z^0 = Z_1^0, \dots, Z_k^0$ , and set  $t = 0$ .
- With fixed cluster centers  $Z^t$ , solve  $P(U, Z^t)$  to obtain partition matrix  $U^t$ . Each object  $X_i$  is assigned to the cluster with the nearest cluster center.
- With fixed partition matrix  $U^t$ , generate updated cluster centers  $Z^{t+1}$  to minimize  $P(U^t, Z^{t+1})$ . The new cluster centers are computed as the mean of the objects within each cluster.
- If convergence is reached or a stopping criterion is satisfied, output the final result and terminate. Otherwise, increment  $t$  by 1 and go back to step 2.

By iteratively updating the partition matrix and the cluster centers, the k-means algorithm converges to a local minimum, providing an effective approach for clustering numerical data objects. As mentioned above, the K-means algorithm [3] is primarily designed for numerical data. However, the algorithm's direct application to categorical data presents challenges due to the need for a natural numerical representation for categorical variables. Several extensions and adaptations of the K-means algorithm have been proposed to overcome this limitation to enable its use with categorical data. One widely used extension is the K-modes algorithm [3]. Unlike the original K-means algorithm, which relies on the Euclidean distance, the K-modes algorithm utilizes a dissimilarity measure specifically tailored for categorical variables. Instead of computing distances based on coordinates in a multidimensional space, the K-modes algorithm uses a simple matching distance measure and defines "cluster centers" as modes. In the K-modes algorithm, the dissimilarity between two categorical objects,  $X$  and  $Y$  described by  $M$  categorical attributes, is computed by counting the total number of matching attribute values between the two objects. The dissimilarity measure is defined as shown in Eq. (3):

$$dis(X, Y) = \sum_{i=1}^M \delta(X_i, Y_i) \quad (3)$$

where,

$$\delta(X_i, Y_i) = \begin{cases} 0 & \text{if } X_i = Y_i \\ 1 & \text{if } X_i \neq Y_i \end{cases}$$

For a cluster of categorical objects,  $\{X_1, \dots, X_N\}$ , where  $X_i = (x_{i1}, \dots, x_{iM})$  and  $1 \leq i \leq N$ , the K-modes algorithm defines the mode  $Z = (o_1, \dots, o_M)$  of the cluster by assigning  $o_m, 1 \leq m \leq M$ , as the most frequently appearing value within  $\{x_{1m}, \dots, x_{Nm}\}$ . The authors in [3] introduced these modifications to develop the K-modes algorithm, which resembles the K-means method for clustering categorical data. However, it should be noted that the mode of a cluster is not generally unique, which introduces instability into the algorithm depending on the selection of modes during the clustering process.

The k-Representative algorithm [7] is a further extension of the K-means algorithm which incorporates the idea of cluster representatives. Rather than utilizing modes as cluster centers, this concept was brought forward by [7], defining representatives in the following manner. Let's take a cluster  $S$ , comprised of categorical objects, expressed as  $S = \{X_1, \dots, X_p\}$ . Each object  $X_i$  can be represented as  $(x_{i1}, \dots, x_{iM})$  with the condition  $1 \leq i \leq p$ . For each attribute  $m$  ranging from 1 to  $M$ ,  $O_m^S$  symbolizes the set of categorical values derived from  $x_{1m}, \dots, x_{pm}$  within the cluster, that means  $O_m^S$  denotes the set of unique categorical values for attribute  $m$  within a specific cluster  $S$ . This is essentially a collection of all distinct categories for attribute  $m$  across all objects in the cluster  $S$ . For instance, suppose a cluster  $S$  consisting of the following three objects:

- Object 1: (Red, Circle, Large)
- Object 2: (Blue, Circle, Medium)
- Object 3: (Red, Square, Medium)

Upon examining the attribute 1 (color), then  $O_1^S$  would be {Red, Blue}, since these are the unique color values within cluster  $S$ . Similarly,  $O_2^S$  for attribute 2 (shape) would be {Circle, Square}, and  $O_3^S$  for attribute 3 (size) would be {Large, Medium}. The representative of cluster  $S$ , denoted by  $Z_S = (z_1^S, \dots, z_M^S)$ , is characterized as:

$$z_m^S = \{(o_{ml}, fS(o_{ml})) \mid o_{ml} \text{ is an element of } O_m^S\} \quad (4)$$

Here,  $fS(o_{ml})$  signifies the proportional frequency of category  $o_{ml}$  within cluster  $S$ . This is computed by dividing the number of objects in  $S$ , possessing the category  $o_{ml}$  for the  $m^{th}$  attribute, by the total count of objects in  $S$ . This is represented by  $\#S(o_{ml})$  and  $p$ , respectively:

$$fS(o_{ml}) = \#S(o_{ml}) / p \quad (5)$$

In essence, each  $z_m^S$  is a distribution over  $O_m^S$ , defined by the proportional frequencies of categorical values inside the cluster. The k-Representative algorithm employs a simple matching measure,  $\delta$ , to determine the dissimilarity between an object  $X = (x_1, \dots, x_M)$  and the representative  $Z_S$ . The dissimilarity  $dis(X, Z_S)$  is characterized as defined in Eq. (6):

$$dis(X, Z_S) = \sum_{m=1}^M \sum_{o_{ml} \in O_m^S} fS(o_{ml}) \cdot \delta(x_m, o_{ml}) \quad (6)$$

In this context, the dissimilarity  $dis(X, Z_S)$  is predominantly influenced by two key factors: the proportional frequencies of categorical values within the cluster and the basic matching of these categorical values. The proportional frequencies reflect the relative importance and prevalence of different categories within the cluster, while the matching mechanism assesses the similarity between the categorical values of the data object  $X$  and the representative center  $Z_S$ . By considering both the proportional frequencies and the matching aspect, the dissimilarity measure captures the distinctive characteristics and relationships of categorical variables within the cluster. To illustrate this, let's continue with the example above and consider a new object 4: (Blue, Circle, Small). The dissimilarity between object 4 and the representative  $Z_S$  of the cluster  $S$  would be calculated as follows: The representative  $Z_S$  for the previous example cluster would be:

- For attribute 1 (Color): {'Red', 0.67}, ('Blue', 0.33)}
- For attribute 2 (Shape): {'Circle', 0.67}, ('Square', 0.33)}
- For attribute 3 (Size): {'Large', 0.33}, ('Medium', 0.67)}

Let's calculate the dissimilarity between Object 4 and the representative  $Z_S$  using Eq. (6). For each attribute, the contribution to the overall dissimilarity is individually calculated:

For attribute 1 (Color):

$$o_{ml} \text{ in } O_1^S: \{('Red'), ('Blue')\}$$

$$fS(o_{ml}): \{0.67, 0.33\}$$

$$\delta('Blue', 'Red')=1, \quad \delta('Blue', 'Blue')=0$$

Contribution for attribute 1:

$$fS('Red') \cdot \delta('Blue', 'Red') + fS('Blue') \cdot \delta('Blue', 'Blue') \\ = 0.67 \cdot 1 + 0.33 \cdot 0 = 0.67$$

For attribute 2 (Shape):

$$o_{ml} \text{ in } O_2^S: \{('Circle'), ('Square')\}$$

$$fS(o_{ml}): \{0.67, 0.33\}$$

$$\delta('Circle', 'Circle')=0, \quad \delta('Circle', 'Square')=1$$

Contribution for attribute 2:

$$fS('Circle') \cdot \delta('Circle', 'Circle') + fS('Square') \cdot \delta('Circle', 'Square') \\ = 0.67 \cdot 0 + 0.33 \cdot 1 = 0.33$$

For attribute 3 (Size):

$$o_{ml} \text{ in } O_3^S: \{('Large'), ('Medium')\}$$

$$fS(o_{ml}): \{0.33, 0.67\}$$

$$\delta('Small', 'Large')=1, \quad \delta('Small', 'Medium')=1$$

Contribution for attribute 2:

$$fS('Large') \cdot \delta('Small', 'Large') + fS('Medium') \cdot \delta('Small', 'Medium') \\ = 0.33 \cdot 1 + 0.67 \cdot 1 = 1$$

Finally, sum up the contributions from all attributes:

$$dis(Object4, V_S) = 0.67 + 0.33 + 1 = 2$$

Therefore, the dissimilarity between object 4 and the representative  $Z_C$  of cluster  $S$  is 2. Based on this dissimilarity, the cluster assignment of object 4 can be determined by comparing its dissimilarity with other cluster representatives. Object 4 will be assigned to the cluster with the lowest dissimilarity, indicating the cluster it is most similar to.

Several extensions have been proposed to tackle specific issues related to categorical data clustering. The k-Centers algorithm [8] is an extension of the K-means algorithm, defining the cluster center as a set of probability distributions estimated using a kernel density estimation method. Dissimilarities between data objects and cluster centers are calculated using indicator vectors and squared Euclidean distance, providing an effective method for clustering categorical data while preserving the key principles of K-means. Other extensions have been proposed to address specific challenges associated with clustering categorical data. These extensions include fuzzy K-modes [20], scalable K-modes [21], and probabilistic K-modes [22], among others. Fuzzy K-modes permit soft assignments, allowing data points to belong to multiple clusters with various degrees of membership. Scalable K-modes enhance computational efficiency for large-scale categorical datasets, while probabilistic K-modes incorporate probabilistic models for uncertainty in cluster assignments, giving a more refined view of cluster membership.

The development of these extensions demonstrates the ongoing efforts to adapt the K-means algorithm for categorical data clustering. These approaches not only consider the unique characteristics of categorical variables but also address challenges such as missing data, scalability, and uncertainty in cluster assignments. The availability of these extensions expands the applicability of the K-means algorithm, allowing it to be effectively utilized in a wide range of domains where categorical data analysis is prevalent.

## V. PROPOSED CLUSTERING METHODS

### A. K-means Dijkstra on Lattice (KDL)

The K-means Dijkstra on Lattice (KDL) method uniquely merges the structural representation of Formal Concept Analysis (FCA) with the computational efficiency of a customized version of Dijkstra's algorithm. This innovative procedure addresses the distinctive challenges associated with clustering categorical data while duly considering the data's inherent structure. KDL method harmoniously integrates the mathematical rigor of FCA and the algorithmic strength of Dijkstra's approach, crafting a new path in categorical data analysis. The general procedure of KDL as follows:

- **Data Conversion to Formal Context:** In the initial phase, the categorical data is transformed into a formal context. This context is represented by a binary matrix, where rows correspond to distinct objects, and columns represent various attributes. A value of 1 signifies that a given object belongs to a certain attribute category, and a 0 indicates the lack of such a relationship.
- **Formal Concept Derivation:** The FCA is then applied to this formal context. It uses the "NextClosure" algorithm

[23], which ensures that all possible formal concepts in the context are generated. These concepts represent significant associations between objects and attributes and capture the underlying structure and dependencies within the data. The hierarchical relationships among these concepts are represented in a lattice structure or graph, which is constructed using the "Ipred" algorithm, modified to suit the needs of the current study. "Ipred" algorithm is very fast for building the Hasse diagram. For more information about this algorithm, refer to [24]. In the quest for an analytical solution to count concepts, Schüt [25] proposed an upper approximation for the count, expressed as  $|C| \leq 3/2 \cdot 2^{\sqrt{|I|+1}} - 1$ , where  $|C|$  denotes the number of concepts and  $|I|$  represents the number of entries in the formal context. This approximation accommodates not merely the number of objects or attributes but also the overall size of the context, providing a potentially more precise estimate of the concept count [26].

- **Assigning Edge Weights:** This step is critical for defining the cost of moving between concepts within the lattice. Each transition has an associated cost, which can vary depending on the direction of movement. For example, transitioning from a parent concept to a child concept can be assigned a higher cost (let's say a cost of 2), than moving from child to parent (let's say a cost of 1), reflecting the importance of specific transitions in categorical attributes.
- **Using Dijkstra's Algorithm for Distance Measurement:** The method employs the Dijkstra algorithm to measure distance within the concept lattice. It calculates the minimum cost of the shortest path between two formal concepts in a concept lattice using the weights assigned to the edges. This provides an effective measurement for the optimal path within the concept lattice.
- **Calculation and Updating of Cluster Centroids:** The centroids of the clusters, which need to be formal concepts themselves, are calculated and continually updated until the values stabilize. The representative centroid of a cluster is the formal concept with the smallest sum of distances to all other concepts within the cluster, minimizing the overall clustering cost function.

In the context of the proposed clustering method utilizing Formal Concept Analysis (FCA) and Dijkstra's algorithm, an important property arises, for any pair of concepts  $c_1$  and  $c_2$  within the concept lattice, there always exists a path connecting  $c_1$  to  $c_2$ . The concept lattice, constructed through FCA, represents the hierarchy of formal concepts derived from the categorical data. The property asserts that no matter which two concepts are chosen within the lattice, there is always a path connecting them. This means there is a sequence of edges to traverse, moving from one concept to another, ultimately leading from  $c_1$  to  $c_2$ . By relying on the subconcept-super concept relation ( $\leq$ ) transitivity in the lattice prove that a path exists between any two concepts in a lattice as described in Definition 4 in Section II. Let's consider two concepts  $c_1$  and  $c_2$  in the lattice. If  $c_1$  and  $c_2$  are directly connected (i.e.,  $c_1 \leq c_2$  or

$c_2 \leq c_1$ ), then a path exists between them as they are adjacent concepts in the lattice. If  $c_1$  and  $c_2$  are not directly connected, consideration can be given to all concepts that are reachable from  $c_1$  in the lattice. Let's denote this set as  $R(c_1)$ . Similarly, the set of all concepts reachable from  $c_2$  as can be defined  $R(c_2)$ . Since the lattice is a partially ordered set,  $R(c_1)$  and  $R(c_2)$  are subsets of the lattice. Now, let's consider the intersection of  $R(c_1)$  and  $R(c_2)$ , denoted as  $R(c_1) \cap R(c_2)$ . If their intersection is not empty, it means that there exists at least one concept that is reachable from both  $c_1$  and  $c_2$ . Let's denote this concept as  $c$ . Since  $c$  is in  $R(c_1)$ , there is a path from  $c_1$  to  $c$ . Similarly, because  $c$  is within  $R(c_2)$ , a path leads from  $c$  to  $c_2$ . Thus, by connecting these two paths, a continuous path is formed from  $c_1$  to  $c_2$ .

If  $R(c_1) \cap R(c_2)$  is empty, no concepts can be reached from both  $c_1$  and  $c_2$ . Nonetheless, establish a pathway between  $c_1$  and  $c_2$  by considering the concepts that are reachable from each and locating a shared concept that functions as an intermediary. This process can be conducted iteratively, broadening the search for concepts reachable from both  $c_1$  and  $c_2$  until a common concept is discovered. By exploiting the transitivity of the subconcept-super concept relationship and considering the accessibility of concepts within the lattice, a path between any two lattice concepts can always be found.

The concept lattice structure inherently guarantees a path between any pair of concepts, rooted in its construction, which encapsulates all potential combinations of objects and attributes as formal concepts. As a result, the lattice forms a connected structure, allowing for traversal from one concept to another through a series of links. This trait is vital for the suggested clustering method, ensuring the computation of the least costly shortest path between any formal concept pair using the Dijkstra-based distance measure. It confirms that all concepts in the lattice can be accessed, and their dissimilarities compared, facilitating precise cluster assignments based on categorical profiles. The clustering method efficiently leverages this connectivity within the concept lattice, capturing hierarchical relationships, semantic constraints, and directional costs embedded in the data. This inherent lattice connectivity aids in delivering meaningful and accurate cluster assignments, ensuring all concepts within the lattice remain interconnected and accessible.

1) *Dijkstra-based distance measure description:* The Dijkstra-Based Distance Measure is a key component of the K-means Dijkstra on Lattice (KDL) method, serving as a more efficient substitute for the Euclidean distance metric in standard K-means algorithms. The KDL method incorporates Dijkstra's algorithm on a lattice structure generated from categorical data via Formal Concept Analysis (FCA). The algorithm computes the shortest path between two formal concepts in the lattice, considering both cost and path direction. Notably, upward transitions (parent-to-child concept) may incur higher costs than downward transitions (child-to-parent). The algorithm utilizes a min-heap-based priority queue for optimization.

Formally, the Hasse diagram built from a concept lattice  $B(C, <)$ , can be represented as  $\mathcal{H}(C, E)$ , where  $C$  is the set of

formal concepts, and  $E$  denotes the edges representing the hierarchical relationships among them. The start and end formal concepts are denoted as  $C_s$  and  $C_e$ , respectively. The cost to reach a concept  $c$  from the start concept  $C_s$  is represented as  $d(c)$ . The algorithm also defines cost functions "UpCost" and "DownCost," representing the costs of moving upwards and downwards in the lattice. The Dijkstra-based distance measure uses a priority queue  $Q$ , based on a min-heap, where each element is a pair  $(d(c), c)$  sorted by  $d(c)$ . It also maintains a set  $V$  to keep track of the nodes already visited. The cost function  $f: C \times C \rightarrow \mathbb{R} \cup \{\infty\}$  is defined with  $c$  and  $c'$  being two formal concepts in the lattice:

$$f(c, c') = \begin{cases} \text{UpCost}, & \text{if } c \supseteq c' \\ \text{DownCost}, & \text{otherwise} \end{cases} \quad (7)$$

Subsequently, the Dijkstra-based distance measure, represented as  $d: C \times C \rightarrow \mathbb{R} \cup \{\infty\}$ , computes the minimum cost path from the start concept  $C_s$  to the end concept  $C_e$  in the lattice:

$$d(C_s, C_e) = \min\{\sum_{i=1}^{n-1} f(c_i, c_{i+1}) \mid (c_1, c_2, \dots, c_n) \text{ is a path from } C_s \text{ to } C_e\} \quad (8)$$

Here,  $n$  is the number of formal concepts in a specific path from  $C_s$  to  $C_e$ . For each possible path from  $C_s$  to  $C_e$ , sum the costs from each step  $c_i$  to  $c_{i+1}$  in the path, and  $d(C_s, C_e)$  is the minimum of these sums over all possible paths. The algorithm functions as follows:

---

**Algorithm 1:** The Dijkstra-based distance measure algorithm on the concept lattice.

---

**Inputs:**  $C_s, C_e, \mathcal{H}(C, E), \text{UpCost}, \text{DownCost}$ .

**Output:** minimum cost from  $C_s$  to  $C_e$

Initialize:

For each  $c$  in  $\mathcal{H}$ :

$d(c) \leftarrow \infty$

EndFor

$d(C_s) \leftarrow 0$

Initialize  $P$  and  $V \leftarrow \emptyset$

Insert  $(0, C_s)$  into  $Q$

While  $Q \neq \emptyset$  do:

$(d(c), c) \leftarrow \text{Dequeue}(Q)$

If  $c = C_e$  then:

Return  $d(C_e)$ .

EndIf

If  $c$  not in  $V$  then:

Add  $c$  to  $V$

EndIf

For each neighbor  $u$  of  $c$  do:

If neighbor not in  $V$ :

If  $c$  is a superset of  $u$  then:

$cost \leftarrow d(c) + \text{UpCost}$

Else:

$cost \leftarrow d(c) + \text{DownCost}$

EndIf

If  $cost < d(u)$  then:

$d(u) \leftarrow cost$

$P(u) \leftarrow c$

Enqueue  $(d(u), u)$  into  $Q$

EndIf

EndIf

EndFor  
EndWhile

---

In Algorithm 1, several symbols are introduced for clarity. The symbol  $C_s$  denotes the starting concept in the concept lattice, while  $C_e$  represents the ending concept.  $\mathcal{H}(C, E)$  refers to the concept lattice itself, composed of concepts  $C$  and edges  $E$ . The terms 'UpCost' and 'DownCost' specify the costs for upward and downward movements within the lattice, respectively.  $d(c)$  is used to signify the current shortest path distance from  $C$  to any given concept  $c$ . A predecessor map is denoted by  $P$ , where  $P(c)$  reveals the predecessor of a concept  $c$  in the shortest path originating from  $C_s$ . Finally,  $Q$  and  $V$  serve as a priority queue for upcoming nodes to visit and a set for nodes already visited, respectively. The algorithm always returns the minimum cost of the shortest path between  $C_s$  and  $C_e$  due to the property of the lattice structure, which ensures a path exists between any two concepts.

This approach has a time complexity of  $O(E + C \log(C))$ , where  $E$  is the number of edges (relationships between formal concepts), and  $C$  is the total count of formal concepts in the lattice. By leveraging the lattice structure, the cost function, and an efficient min-heap-based priority queue, the Dijkstra-Based Distance Measure provides a more accurate representation of dissimilarities in categorical data. This results in an optimized clustering process and yields more accurate and meaningful cluster assignments.

2) *Cluster centers:* Defining the cluster centers, or centroids, in a concept lattice is vital for effectively implementing the K-means Dijkstra on Lattice (KDL) method. These centroids need to be formal concepts within the lattice. The continual updating and calculation of these representative centroids significantly influence the minimization of the overall clustering cost function. To formally describe this, consider a cluster  $S$  composed of a set of formal concepts  $\{c_i\}$  where  $i = 1, 2, \dots, |S|$ . The representative formal concept, denoted as  $Z$ , is defined as the concept within  $S$  that minimizes the sum of the distances to all other concepts in the same cluster. This can be mathematically expressed as:

$$Z = \operatorname{argmin}_{Z \in S} \left( \sum_{i=1}^{|S|} d(c_i, Z) \right) \quad (9)$$

Here,

$c_i$  represents each concept within the cluster  $S$

In Eq. (9),  $d(c_i, Z)$  is the Dijkstra-based distance from the potential centroid  $Z$  to each concept  $c_i$  within the cluster. The argmin operation is employed to find the formal concept  $Z$  in  $S$  that yields the smallest sum of distances to all other concepts in the cluster  $S$ . It is important to note that  $Z$  inherently belongs to the cluster  $S$ , which allows for a more efficient calculation of the minimal sum of distances to all other concepts within  $S$ . Furthermore, the existence of a center for any set of formal concepts is ensured due to the properties of the Dijkstra-based distance measure. This consistency makes the method universally applicable, regardless of the specific set of formal concepts under consideration. This method of defining cluster centers in the concept lattice adheres to mathematical rigor while being practically feasible, offering a systematic way to



manage and interpret complex categorical datasets. This method enhances the interpretability of clustering results by identifying representative formal concepts for each cluster, fostering more comprehensive and insightful data analysis.

3) *The clustering algorithm:* The K-Means Dijkstra on Lattice (KDL) clustering algorithm, grounded in Formal Concept Analysis (FCA) and the Dijkstra-based distance measure, can be articulated through the following sequential stages:

---

**Algorithm 2:** K-Means Dijkstra on Lattice (KDL) clustering algorithm

---

**Inputs:**  $k$ , the number of clusters;  $\mathcal{B}$ , the lattice of formal concepts.

**Output:** The resulting clusters  $\{S_1, S_2, \dots, S_k\}$ .

**Initialize:**

Select  $k$  formal concepts  $\{c_1, c_2, \dots, c_k\}$  from the lattice  $\mathcal{B}$  randomly as the initial centroids of the  $k$  clusters.

**Assignment:**

For each formal concept  $c \in \mathcal{B}$  do:

Assign  $c$  to the cluster  $S_i$  for which the Dijkstra-based distance measure  $d(c, Z_i)$  is minimized, where  $Z_i$  is the centroid of cluster  $S_i$ .

Using Equations (7, 8)

**Centroid Update:**

For each cluster  $S_i$  do:

Recalculate the centroid  $Z_i$  as the formal concept  $c$  that minimizes the total distance to all other concepts within  $S_i$

Using Equation (9)

**Iteration:**

While centroids change between iterations do:

Repeat steps 2 and 3.

**Finalization:**

Output the resulting clusters  $\{S_1, S_2, \dots, S_k\}$ .

---

4) *Cost analysis of k-means dijkstra on lattice (KDL) method:* This section analyzes the computational complexity of the proposed K-Means Dijkstra on Lattice (KDL) method. The computational cost of each step will be evaluated, from the initialization of clusters to their final assignment. This analysis will provide insights into the efficiency and scalability of the KDL method.

Given the parameters:

- $K$  represents the number of clusters.
- $N$  represents the number of objects.
- $A$  represents the number of attributes.
- $C$  represents the number of concepts.
- $E$  represents the number of edges in the lattice.
- $B$  represents the maximum number of border elements in the lattice construction.

The proposed K-Means Dijkstra on Lattice (KDL) method commences with data preprocessing, wherein the categorical dataset transforms a formal context. This stage involves a binary translation of each dataset entry, leading to a time complexity of  $O(NA)$ . Subsequently, a lattice is generated from the formal concepts obtained in the previous stage. This

phase necessitates looping over all border elements for each concept, inducing a worst-case time complexity of  $O(CB)$ . The final stage of the process encompasses a K-means-like clustering operation. This phase includes iterations over all the concepts to assign clusters and update centroids. The computation of the shortest paths between pairs of concepts within the lattice primarily influences this stage's time complexity. This is achieved using Dijkstra's distance algorithm. Assuming  $I$  iterations are required to reach convergence, the time complexity for computing the shortest paths between all pairs of concepts culminates in  $O(ICK(E + C \log C))$ . To summarize, the KDL method's overall time complexity, significantly influenced by the data preprocessing, lattice construction, and lattice-based clustering stages, can be approximated as  $O(NA + CB + ICK(E + C \log C))$ . It is worth noting that this is a rough estimation, with actual time complexity potentially varying based on the characteristics and data distribution within the formal context. However, focusing on the dominant term for the sake of simplification, the time complexity of the KDL method becomes  $O(ICK(E + C \log C))$ .

**B. K-means Vector on Lattice (KVL)**

The K-means Vector on Lattice (KVL) method is essential for converting categorical data into numerical data. Leveraging the classical k-means algorithm facilitates data grouping, making it instrumental for various data analysis operations, especially when dealing with predominantly categorical or non-numeric data. The essence of this method lies in its capacity to convert formal concepts, regardless of their abstract or categorical nature, into 'concept description vectors'. These vectors exist in a real-valued vector space, which not only makes them easily adaptable to standard mathematical procedures but also optimizes them for computational analysis. Each vector represents the formal concept from which it was derived, encapsulating its fundamental attributes. Every dimension within the vector signifies a different attribute of the concept, with its magnitude corresponding to the prevalence or significance of the attribute within the concept. This forms a compressed yet efficient way to contain the information intrinsic to the formal concept.

With the creation of concept description vectors, these entities can now be subjected to the k-means algorithm. This renowned clustering method partitions the data into a specified number 'k' of distinct clusters. Each cluster is identified by its centroid, which serves as the symbolic or physical center of the cluster. All data points, or concept description vectors in this context, within a specific cluster have a closer similarity to their cluster's centroid than to those of other clusters. This aids in the aggregation of analogous concepts, thereby facilitating insightful analysis of the data.

**Definition 5. Concept Description Vector:** Let  $Y = (A, B)$  be a formal concept, where  $A \subseteq G$ ,  $B \subseteq M$ , and a context  $T = (G, M, I)$  has  $|M| = q$ ,  $|G| = r$ , the incidence relation  $I \subseteq G \times M$  can be represented as a binary matrix, where the rows correspond to the elements of  $G$  (objects), the columns correspond to the elements of  $M$  (attributes), and each entry of the matrix is either 1 or 0, indicating whether the relation  $(g, m)$  exists or not. Let's denote this matrix as  $I$ , with

dimensions  $r \times q$ , where  $r$  is the number of objects in  $G$  and  $q$  is the number of attributes in  $M$ . The rows of the matrix can be labeled by  $g_1, g_2, \dots, g_r$  and the columns by  $m_1, m_2, \dots, m_q$ , the matrix can be defined as shown in Table II:

The concept description vector is defined as  $V_Y = (v_{m_1}, v_{m_2}, \dots, v_{m_q})$ .  $v_{m_h}$  ( $h$  ranging from 1 to  $q$ ) is obtained as follows [5]:

$$v_{m_h} = \begin{cases} 1 & \text{if } m_h \in B \\ \frac{1}{r} \sum_{j=1}^r I(g_j, m_h) & \text{if } m_h \notin B, \forall g_j \in G \end{cases} \quad (10)$$

where  $m_h \in M$ .

The method to calculate each attribute within a concept is different. It depends on whether the attribute is in the intent of each concept or not. The concept vector is the base for getting the similarity between concepts. This vector is obtained from the context based on the intent of each concept. Depending on whether the attribute  $m_h$  is a part of the intent  $B$ . If  $m_h$  is a part of the intent, it is assigned a value of 1, indicating its high importance in defining the concept. If not, it's calculated as the average association of this attribute across all objects, represented as  $\frac{1}{r} \sum_{j=1}^r I(g_j, m_h)$  if  $m_h \notin B$ . This can be perceived as the frequency or relevance of this attribute across the object set  $G$ . The computation of each attribute forms the concept description vector used to ascertain the similarity between concepts.

After defining the concept description vectors, the KVL method introduces the concept similarity measure. This measure, often referred to as Concept Similarity ( $CS$ ), as explicated in Definition 6, is used to ascertain the proximity of these concepts. Concept Similarity is calculated using the Euclidean Distance between any two concept description vectors  $V_{Y_1}$  and  $V_{Y_2}$ . The  $CS$  equation helps us to quantify the closeness between two concepts, taking into account each component of their respective concept description vectors.

**Definition 6. Concept Similarity:** Concept Similarity ( $CS$ ) is calculated based on the concept description vector in Definition 5 using traditional Euclidean distance. For any two concept description vectors

$V_{Y_1} = (V_{Y_1 m_1}, V_{Y_1 m_2}, \dots, V_{Y_1 m_q})$  and  $V_{Y_2} = (V_{Y_2 m_1}, V_{Y_2 m_2}, \dots, V_{Y_2 m_q})$ , the Euclidean distance is defined as per Eq. (5):

$$CS(V_{Y_1}, V_{Y_2}) = \frac{1}{\sqrt{(V_{Y_1 m_1} - V_{Y_2 m_1})^2 + (V_{Y_1 m_2} - V_{Y_2 m_2})^2 + \dots + (V_{Y_1 m_q} - V_{Y_2 m_q})^2}} \quad (11)$$

TABLE II. MATRIX CORRESPONDING TO THE RELATION  $I$

Objects/Attributes	$m_1$	$m_2$	...	$m_q$
$g_1$	$I(g_1, m_1)$	$I(g_1, m_2)$	...	$I(g_1, m_q)$
$g_2$	$I(g_2, m_1)$	$I(g_2, m_2)$	...	$I(g_2, m_q)$
...	...	...	...	...
$g_r$	$I(g_r, m_1)$	$I(g_r, m_2)$	...	$I(g_r, m_q)$

This framework of concept description vectors and concept similarity lays the groundwork for the k-means clustering algorithm. The algorithm takes the concept description vectors as inputs and leverages the concept similarity measure to identify which concepts most resemble each other. Concepts exhibiting high similarity are then grouped into clusters. The center of each cluster, represented by  $Z_i$ , is calculated as the mean of all concept description vectors within that cluster. Let's denote  $S_i$  as the  $i^{th}$  cluster ( $i = 1, 2, \dots, k$ ), where  $k$  is the number of clusters. The centroid of each cluster  $S_i$  can be defined as:

$$Z_i = \frac{1}{|S_i|} \sum_{j=1}^{|S_i|} V_{Y_j}, V_{Y_j} \in S_i \quad (12)$$

The k-means algorithm aims to minimize the within-cluster sum of squares (WCSS) of Euclidean distances. This objective function  $Q$  is as follows:

$$Q = \sum_{i=1}^k \sum_{j=1}^{|S_i|} \|V_{Y_j} - Z_i\|^2, V_{Y_j}, Z_i \in S_i \quad (13)$$

The algorithm alternates between assigning each concept description vector to the nearest centroid and recalculating the centroid of each cluster using Eq. (11) and (12), until the clusters stabilize. The algorithm has converged, and the clusters are optimally partitioned concerning the given concept description vectors.

**1) The clustering algorithm:** The main idea is as follows. Suppose  $T = (G, M, I)$  is a formal context and  $V(T)$  is the set of all concept description vectors and the number of clusters is  $K$ . Firstly, the initial centers,  $Z_t^0 = (A_t, B_t)$  ( $t = 1, 2, \dots, K$ ), of  $K$  clusters are selected randomly, and the corresponding clusters are  $S_t^0 = \{Z_t^0\}$ . Secondly, join a concept description vector  $v \in V(T)$  into one cluster according to the following rule: if the distance between  $v$  and  $Z_t^0$  is lower than that of  $v$  and other centers, then,  $v$  is put into  $S_t^0$ . Each vector in  $V(T)$  can be adjusted according to this rule. Thirdly, the new center of each cluster can be determined by calculating the mean value of vectors within each cluster and set it as a new center. Finally, repeat the above process till the twice computation of each cluster and center are the same. The algorithm steps are as follows:

---

**Algorithm 3.** K-means clustering of concepts.

---

**Input:** All the description vectors of concepts in  $V(T)$ ,  $k$ .

**Output:** The clusters and corresponding centers.

**Initialize:**

Set  $S_1^i \leftarrow \emptyset, S_2^i \leftarrow \emptyset, \dots, S_k^i \leftarrow \emptyset;$

$i \leftarrow 0;$

Select initial center vectors of  $K$  clusters:  $Z_1^i, Z_2^i, \dots, Z_k^i;$

**Assignment:**

For each  $v \in V(T)$  do:

-Find  $t$  such that

$CS(\text{distance})(v, Z_t^i) \leq CS(\text{distance})(v, Z_j^i), (j = 1, 2, \dots, k)$  then,

$v \in S_t^i;$

EndFor

**Centroid Update:**

For each  $S_t^i$  do:

$Z_t^{i+1} = \frac{1}{|S_t^i|} \sum_{s=1}^{|S_t^i|} v_s, v_s \in S_t^i$ , using Equation (12)

---

$$S_t^{i+1} = \{v \in V(T) | CS(v, Z_t^{i+1}) \leq CS(v, Z_j^{i+1})\}$$

using Equation (11)

EndFor

**Convergence Check:**

If  $Z_t^i = Z_t^{i+1}, S_t^i = S_t^{i+1}, t = 1, 2, \dots, K$ , then

Go to step 5.

Else:

$i = i + 1$ ,

Go to Step 2.

**Output:** clusters  $S_1^i, S_2^i, \dots, S_k^i$  and the corresponding centers  $Z_1^i, Z_2^i, \dots, Z_k^i$ .

Algorithm 1 outlines the process for performing K-means clustering on the set of concept description vectors. The input to the algorithm is the set of all concept description vectors, and the output is the resulting clusters and their centers. The algorithm begins by randomly initializing the clusters and selecting the initial centers for the clusters. It then assigns each concept description vector to the cluster whose center it is closest to, based on the concept similarity measure in Eq. (11). It then updates the clusters' centers based on the mean of the concept description vectors within each cluster using Eq. (12) and repeats this process until the clusters stabilize. mapping the vectors back to the original concepts. Once the K-means clustering is done and the clusters are stabilized, the vectors within each cluster can be traced back to their original concepts. The mapping uses the concept description vectors created in the first step.

**Algorithm 4: Mapping vectors to original concepts.**

**Input:** The clusters  $S_1^i, S_2^i, \dots, S_k^i$  and the corresponding centers  $Z_1^i, Z_2^i, \dots, Z_k^i$ .

**Output:** Clusters of original concepts.

**Initialize:**

$NS_1, NS_2, \dots, NS_k \leftarrow \emptyset$ ,

For  $t = 1$  to  $k$  do:

For each description vector in  $S_t^i$  do:

Map back to its original concept and add to  $NS_t$ .

Using the relationship between the concept description vector and the original concept established in Definition 5.

EndFor

**Output:** the new clusters  $NS_1, NS_2, \dots, NS_k$ , each containing the original concepts.

In this way, the numerical data obtained from the K-means clustering is transformed back into categorical data, giving us clusters of similar concepts rather than clusters of similar vectors. This method of approximation and mapping allows for efficient and meaningful clustering of concepts in a context, making it easier to understand and interpret the relationships and similarities between different concepts.

2) *Cost Analysis of the KVL Method:* This section thoroughly reviews the computational complexity of the KVL method to assess its efficiency and scalability. The first stage is data preprocessing, which involves the transformation of a dataset into a formal context. Here, a binary representation of each object with N objects and A attributes is needed, introducing a time complexity of O(NA). After preprocessing, formal concepts are generated and then converted into vectors in an A-dimensional attribute space. For each of the C

concepts, an equivalent vector in the attribute space needs to be calculated, leading to a time complexity of O(AC). The algorithm starts by randomly selecting K centroids from the pool of C concepts, resulting in a time complexity of O(K). Subsequent stages include iterative assignment and update, where each concept is assigned to its nearest centroid, and centroids are updated based on new assignments. This incurs a time complexity of O(CK) per iteration. These processes are performed I times until convergence, resulting in a total time complexity of this stage being O(ICK).

Lastly, the mapped vectors are reconverted to their original formal concepts, which involves a time complexity of O(CK). Summing up the complexities from each phase, the total time complexity of the KVL method is estimated to be O(NA+AC+K+ICK+CK). This is a heuristic estimate, and time complexity could vary based on data distribution and other runtime factors. However, focusing on the dominant term for simplification, the time complexity of the KVL method becomes O(ICK).

## VI. EXPERIMENTAL RESULTS

In this section, the presented experimental results aimed at demonstrating the performance of the Dijkstra-Based Distance Measure, as well as the performance and scalability of the K-means Dijkstra on Lattice (KDL) and K-means Vector on Lattice (KVL) methods. The experiments were conducted on a Mac system equipped with an Apple M1 chip and 8GB of RAM, running Mac OS 13.2.1.

### A. Testing and Evaluation of Dijkstra-based Distance Measure

In the experimental section, the performance of the distance measure based on Dijkstra's algorithm is rigorously evaluated. The testing process involved the following steps:

1) *Random generation of formal contexts:* five formal contexts are randomly generated with varying sizes and densities; the characteristics of these formal contexts are described in Table III. The density parameter in this context refers to the proportion of filled entries (1s) compared to the total number of possible entries in the binary matrix representation of the formal context. It quantifies how much information is present regarding the relationship between objects and attributes. To explain the density parameter, let's consider an example from Table III: Formal Context1 with 600 objects and 125 attributes. The density for Formal Context1 is 0.10, indicating that, on average, each entry in the binary matrix has a 0.10 probability of being filled (assigned a value of 1). A lower density value implies sparser relationships, where fewer objects belong to the given attributes or categories. In contrast, a higher density value indicates denser relationships, where a larger number of objects are associated with the given attributes.

2) In the analysis, four datasets from the UCI Machine Learning repository are meticulously examined. Prior to conducting any experiments, these datasets are transformed into formal contexts, with details outlined in Table III. The

datasets were chosen based on two key criteria: public accessibility and the categorical characteristics of their attributes. The selected datasets include:

- The Balance-Scale dataset is designed to replicate psychological experiment results. Each instance in this dataset can be labeled based on whether the balance scale leans to the left, right, or is balanced.
- The Breast Cancer dataset obtained from the University of Wisconsin hospitals, classifies each instance into one of two potential categories: benign or malignant.
- The Car Evaluation dataset, which results from a simple hierarchical decision model initially designed for DEX's demonstration, classifies each instance into one of four classes: unacc, acc, good, and vgood.
- Tae dataset representing teaching performance assessment over five semesters (three regular and two summers) includes 151 teaching assistant assignments from the University of Wisconsin-Madison's Statistics Department. All instances fall into one of three categories: low, medium, and high.

3) *Extraction of formal concepts:* The NextClosure algorithm was used to extract the set of formal concepts from each formal context. The number of formal concepts generated from each formal context is shown in Table IV.

4) *Hasse diagram construction:* Utilizing the Ipred algorithm, a Hasse diagram was structured optimally for the case study, as elaborated in Section V. The characteristics of the generated Hasse diagrams are shown in Table V, by considering the density parameter in the construction of the Hasse diagram. The density parameter influences the number of edges and nodes in the Hasse diagram. A denser formal context with a higher density value tends to result in a larger number of formal concepts and, consequently, a more extensive concept lattice with a higher number of edges connecting the concepts. On the other hand, a sparser formal context with a lower density value leads to a smaller concept lattice with fewer edges.

TABLE III. CHARACTERISTICS OF RANDOM AND REAL-WORLD FORMAL CONTEXTS

Formal Contexts	#objects	#attributes	density
Formal Context1	600	125	0.10
Formal Context2	11000	30	0.10
Formal Context3	1350	120	0.05
Formal Context4	2000	20	0.15
Formal Context5	12000	20	0.23
Balance-Scale	625	20	0.20
Breast Cancer	182	35	0.25
Tae	151	101	0.04

Car Evaluation	1728	21	0.28
----------------	------	----	------

TABLE IV. FORMAL CONCEPTS GENERATED FROM THE FORMAL CONTEXTS IN TABLE III

Formal Contexts	#formal concepts.
Formal Context1	29926
Formal Context2	15117
Formal Context3	9882
Formal Context4	2989
Formal Context5	39931
Balance-Scale	1297
Breast Cancer	2569
Tae	276
Car Evaluation	8001

TABLE V. HASSE DIAGRAM TRAITS VIA IPRED ALGORITHM

Formal Contexts	#formal concepts	Inclusion relationship between concepts (edges)
Concept lattice1	29926	122839
Concept lattice2	15117	67040
Concept lattice3	9882	36797
Concept lattice4	2989	12175
Concept lattice5	39931	228427
Balance-Scale	1297	4945
Breast Cancer	2569	9513
Tae	276	619
Car Evaluation	8001	38928

The analysis involved running a Dijkstra-based distance measure on concept lattices generated from five random formal contexts and four real-world datasets. The formal contexts varied in number of objects, attributes, and density. On the other hand, the real-world datasets were diverse, encompassing balance scale, breast cancer, teaching assistant evaluation, and car evaluation data. After generating the formal contexts and preparing the datasets, Formal Concept Analysis (FCA) using the NextClosure algorithm has been performed to derive formal concepts. The count of these formal concepts varied significantly across the contexts and datasets, ranging from as low as 2989 in Formal Context 4 to as high as 39931 in Formal Context 5. Using these formal concepts, Hasse diagrams (concept lattices) constructed with the help of the Ipred algorithm. The concept lattices illustrated the inclusion relationship between concepts. Again, the number of inclusion relationships was directly related to the complexity and size of the corresponding formal context. Subsequently, the distance measure was evaluated. Concept pairs were randomly chosen from each concept lattice, constituting 25% of the total concepts. The minimum cost of the shortest path between these pairs was calculated using the designated distance measure. This evaluation was performed across ten trials, with both the average runtime and mean distance documented for each.

We're observing a comparison of the average run time of the Dijkstra-based distance measure algorithm and the mean distance between concepts for both randomly generated formal datasets and real-world datasets. As indicated in Fig. 2 and

Fig. 3 for randomly generated datasets, there is a clear correlation between the number of concepts in the lattice and the algorithm's runtime. An increase in the number of concepts leads to a corresponding rise in runtime. This finding aligns with expectations, as a lattice with a greater number of concepts and relationships is likely to be more complex. Consequently, calculating the shortest path between pairs in this intricate structure would naturally demand more computational time and resources. The mean distances were mostly consistent for each context, indicating a relatively stable distance measure despite potential variance in the random generation of the formal contexts. This reinforces the efficacy of the Dijkstra-based distance measure, highlighting its stability across multiple trials of randomly generated data.

Patterns similar to those observed in randomly generated datasets were also evident in real-world datasets, as demonstrated in Fig. 4 and Fig. 5. The runtimes correlate with the number of concepts in the lattice, with larger lattices taking longer to calculate the shortest paths. Interestingly, the Car Evaluation dataset, which had the highest number of concepts (8001), exhibited the shortest mean distance (6.5735) among the real-world datasets. This suggests that although the dataset is complex, the relationships within the data are more straightforward or closer than the other datasets. Meanwhile, the Balance-Scale and Breast Cancer datasets had a more moderate number of concepts (1297 and 2569, respectively) and showed a higher mean distance. This might indicate that, despite having fewer concepts, the relationships in these datasets could be more complex or convoluted.

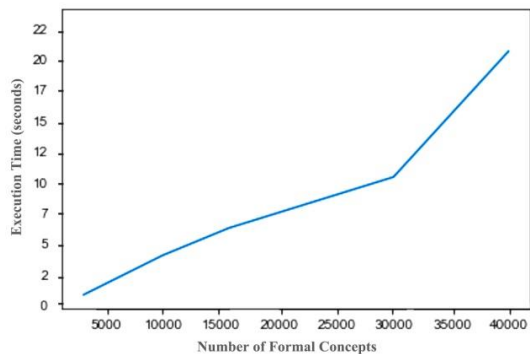


Fig. 2. Average runtime of distance calculation algorithm on different concept lattice sizes for the random contexts in Table IV.

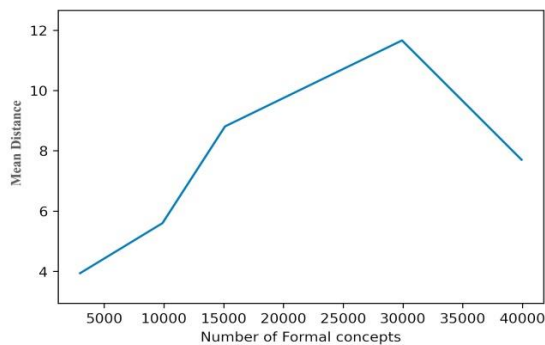


Fig. 3. Mean distance of distance measure algorithm on different concept lattice sizes for the random contexts in Table IV.

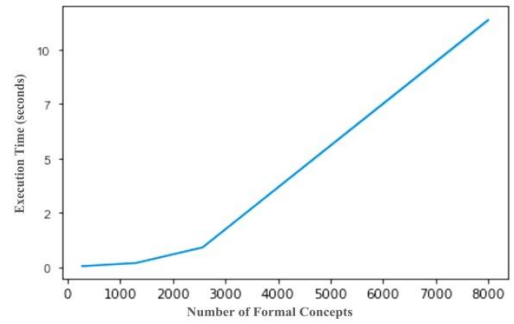


Fig. 4. Average runtime of distance calculation algorithm on different concept lattice sizes of real-world datasets.

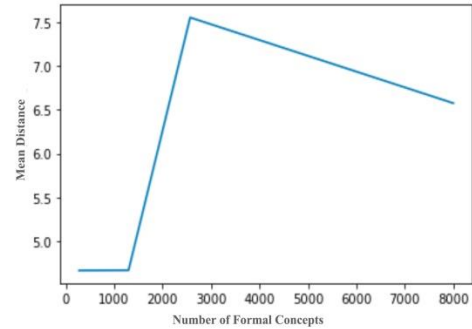


Fig. 5. Mean distance of distance measure algorithm on different concept lattice sizes of real-world datasets.

Overall, the results suggest that the Dijkstra-based distance measure is robust and stable across various randomly generated and real-world contexts. The run time increases as expected with the size and complexity of the dataset, and the measure captures the inherent complexity in the data (as reflected in the mean distances) and provides valuable insights into the structural properties of concept lattices. It allows for identifying concept pairs relatively closer or farther apart within the lattice structure. The results contribute to a better understanding of relationships and structural characteristics within formal contexts and concept lattices. The consistent performance of the measure across different scenarios reinforces its potential utility in handling diverse and complex categorical datasets.

Adapting the FCA, the Dijkstra-based distance measure applies the robust Dijkstra's algorithm to compute the shortest path between two categorical data points. This method, mindful of the hierarchical structure of categorical data, quantifies dissimilarity by evaluating paths within the data space. It provides a viable alternative to Euclidean distance within the clustering context when enhancing the K-means algorithm for categorical data analysis. Replacing Euclidean distance with the Dijkstra-based measure allows the K-means algorithm to cluster categorical datasets better, accurately reflecting the relationships and similarities between categorical variables. Incorporating the Dijkstra-based distance measure in the K-means algorithm aids cluster identification based on categorical patterns, providing meaningful insights and potential applications across numerous domains. It's users in new avenues for examining and interpreting categorical data.

### B. Clustering Performance

In the present section, the performance of two distinct clustering approaches designed for categorical data: K-means Dijkstra on Lattice (KDL) and K-means Vector on Lattice (KVL) is scrutinized. We've chosen the Silhouette Coefficient and the Davies-Bouldin Index (DBI) as the evaluation metrics due to their ability to assess clustering performance without the need for ground truth labels, making them especially useful in real-world applications where such labels might not be available. The Silhouette Coefficient serves as a measure to ascertain the suitability of a data point's allocation to its cluster in comparison to other clusters. The coefficient fluctuates between -1 and 1, with a high positive value implying a well-clustered data point, while a negative one indicates potential misplacement within a cluster. Here's the mathematical expression for the Silhouette Coefficient:

$$\text{Silhouette Score} = (b - a) / \max(a, b) \quad (14)$$

Where 'a' is the mean intra-cluster distance and 'b' is the mean nearest-cluster distance. In contrast, the Davies-Bouldin Index (DBI) is a metric that evaluates the separation and compactness of clusters. A lower DBI value implies an optimal clustering solution. The DBI is calculated as follows:

1) Calculate the average distance between each point in a cluster  $S_i$  and all other points in the same cluster. This is often referred to as the intra-cluster distance. Denote this as  $SC_i$  for cluster  $S_i$ .

$$SC_i = (1 / n_i) \sum ||x - Z_i|| \text{ for } x \in S_i$$

where:

- $n_i$  is the number of points in cluster  $S_i$ ,
- $x$  is a point in cluster  $S_i$ ,
- $Z_i$  is the centroid of cluster  $S_i$ ,
- $||x - Z_i||$  is the distance between point  $x$  and centroid  $Z_i$ .

2) Calculate the distance  $d_{ij}$  between cluster  $S_i$  and  $S_j$ , using a suitable distance measure between the centroids of the clusters.

3) Calculate the ratio  $R_{ij}$  between the sum of the intra-cluster distances of cluster  $S_i$  and  $S_j$ , and the inter-cluster distance between  $S_i$  and  $S_j$ .

$$R_{ij} = (SC_i + SC_j) / d_{ij}$$

4) For each cluster  $S_i$ , find the maximum ratio  $R_i$  which is the maximum  $R_{ij}$  for all  $j \neq i$ .

$$R_i = \max(R_{ij}) \text{ for all } j \neq i$$

5) The Davies-Bouldin Index (DBI) is the average of all  $R_i$ .

$$\Delta \text{BI} = (1 / S) \sum R_i \quad (15)$$

where,  $S$  is the total number of clusters

Again, lower DBI values indicate better clustering because this signifies clusters that are more compact (lower intra-cluster

distances  $SC_i$ ) and better separated (higher inter-cluster distances  $d_{ij}$ ).

The analysis is based on four real-world datasets described in the previous section as shown in Table V and the results in Tables VI and VII. The clustering is performed by setting the number of clusters (k value) equal to the number of classes for each dataset to maintain consistency with the inherent data structure. These results are recorded from the averages of 100 runs for each method. A closer examination of these tables allows for a comparative analysis of the performance of the K-means Dijkstra on Lattice (KDL) and K-means Vector on Lattice (KVL) methods. Regarding the Silhouette Coefficient (Table VI), it is evident that the KDL method, which utilizes the inherent lattice graph structure of categorical data for clustering, consistently outperforms the KVL method, regardless of the dataset used. This outcome is further corroborated by the DBI results (Table VII), where the KDL method again demonstrates superior performance by consistently achieving lower index values across all datasets. This can be attributed to the design of the KDL method. The KDL strategy focuses on integrating the representation of categorical data based on the graph structure, effectively leveraging the potential similarity between these data points. It employs Formal Concept Analysis (FCA), a mathematical framework for generating a concept hierarchy, and Dijkstra's algorithm to calculate the shortest path between formal concepts in the FCA graph. This novel distance measure, which represents the minimal cost of moving from one formal concept to another, facilitates a more accurate clustering process.

On the contrary, the KVL method, while simplifying the clustering process by converting categorical data into numerical vectors and using standard k-means algorithms, risks obscuring the inherent hierarchical relationships between categorical values. This transformation can potentially result in less effective clustering performance.

TABLE VI. SILHOUETTE COEFFICIENT SCORES OF CLUSTERING PERFORMANCE FOR K-MEANS DIJKSTRA ON LATTICE (KDL) AND K-MEANS VECTOR ON LATTICE (KVL) METHODS ACROSS DIVERSE DATASETS

Datasets	KDL	KVL	#Clusters
Balance-Scale	<b>0.406</b>	0.128	3
Breast Cancer	<b>0.239</b>	0.090	2
Tae	<b>0.300</b>	0.092	3
Car Evaluation	<b>0.563</b>	0.106	4

TABLE VII. DBI SCORES OF CLUSTERING PERFORMANCE FOR K-MEANS DIJKSTRA ON LATTICE (KDL) AND K-MEANS VECTOR ON LATTICE (KVL) METHODS ACROSS DIVERSE DATASETS

Datasets	KDL	KVL	# Clusters
Balance-Scale	<b>1.48</b>	2.64	3
Breast Cancer	<b>1.83</b>	2.78	2
Tae	<b>1.49</b>	2.62	3
Car Evaluation	<b>1.90</b>	2.92	4

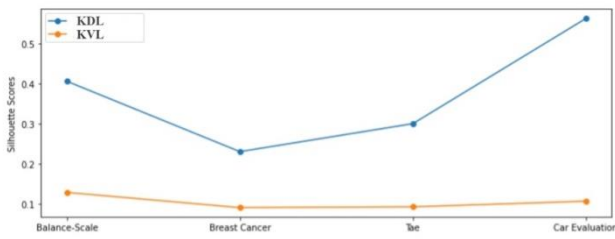


Fig. 6. Silhouette scores by dataset and method.

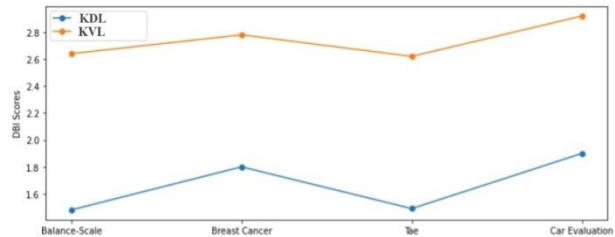


Fig. 7. DBI scores by dataset and method.

The compelling evidence in Tables VI and VII and their corresponding graphical representations in Fig. 6 and Fig. 7 illuminate the KDL method's superior performance over the KVL method for clustering categorical data. By directly handling categorical data and leveraging its inherent hierarchical structure, the KDL method offers more meaningful and accurate clustering results. This finding corroborates the hypothesis that leveraging the inherent similarities and structure of categorical data can yield improved results. While the KVL method simplifies the process by transforming categorical data into numerical vectors, it potentially obscures the intricate hierarchical relationships between categorical values, thus diminishing the method's effectiveness. This is reflected in the lower Silhouette, and higher DBI scores observed for the KVL method. These findings not only present solid evidence in favor of the KDL method as a more potent tool for clustering categorical data, but they also underscore the importance of utilizing the data's inherent structure where possible. However, these conclusions should uphold the utility of the KVL method. Instead, they serve as a critical reminder of the significance of selecting an appropriate tool for the data at hand, considering each method's potential trade-offs and benefits.

### C. Scalability Test Results Analysis

1) Scalability in relation to the number of clusters: To bolster the robustness of the study concerning the K-means Dijkstra on Lattice (KDL) and K-means Vector on Lattice (KVL) clustering techniques, a meticulous analytical approach was employed. This rigorous methodology underscores the credibility of the performance assessments and findings presented. All results presented in this analysis were derived from the average runtime of five independent runs. This method was utilized to mitigate any outliers' influence and deliver a more precise portrayal of each method's performance.

The investigation was particularly interested in the scalability of these methods in response to an increase in the

number of clusters. The number of clusters was varied from 2 to 18 in the analysis, with the dataset size held constant. This aspect is essential in real-world situations, especially when data is complex and needs to segregate into a limited number of clusters neatly. The performance of both methods in relation to the varying number of clusters was assessed using the 'Car Evaluation' dataset consisting of 8001 formal concepts. From Fig. 8, it is evident that the K-means Vector on Lattice (KVL) method demonstrates scalability. A linear relationship is observed between execution time and the increment in the number of clusters. The execution time varies approximately between 44.48 and 51.56 seconds as the number of clusters changes from 2 to 18. This pattern underscores the efficiency of the KVL method in handling larger and more complex datasets.

This method, thus, shows promise in effectively managing a rise in the number of clusters without causing a substantial increase in execution time. On the other hand, Fig. 9 provides insights into the scalability of the KDL method. This method shows rapid growth in execution time as the number of clusters increases. The time jumps from about 1926.77 seconds for 2 clusters to a massive 49600.10 seconds for 18 clusters. Given the complexity of the lattice graph and the number of formal concepts, the KDL method's computational load increases significantly with the number of clusters, suggesting lower scalability.

The KVL method is better in terms of scalability and efficiency for an increasing number of clusters; the KDL method provides higher-quality clustering, though it demands significantly more computational time and resources. This highlights the importance of finding the right balance between computational efficiency and clustering quality. The preference for one over the other may vary depending on the specific situation and constraints.

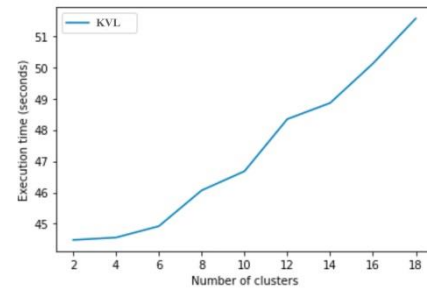


Fig. 8. Scalability of KVL Method to the number of clusters when clustering 8001 formal concepts of the 'car evaluation' dataset.

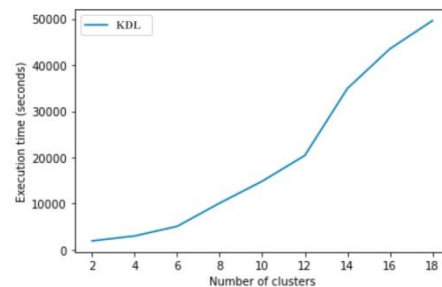


Fig. 9. Scalability of the KDL method to the number of clusters when clustering 8001 formal concepts of the 'car evaluation' dataset.

2) *Scalability in relation to the number of formal concepts*: In examining the scalability of KDL and KVL, performance was assessed with an increasing number of formal concepts, while keeping the cluster count constant at three. This analysis is grounded in multiple iterations of these methods on a selection of real-world datasets, namely Balance-Scale, Breast Cancer, Tae, and Car Evaluation, described in detail in Tables III, IV, and 5. Fig. 10, and Fig. 11, represent the scalability of the KVL and KDL methods, respectively, demonstrating how they fare with a rising count of formal concepts.

Diving into Fig. 10, it's evident that the KVL method shows admirable consistency. The recorded execution times from five separate runs, 43.14, 43.25, 44.15, and 46.35 seconds, corresponding to the datasets featuring 276, 1297, 2569, and 8001 formal concepts. This suggests that as the number of formal concepts increases, the KVL method retains its efficiency, reflecting robust scalability - an attribute crucial for managing large datasets. In comparison, Fig. 11 encapsulates the performance of the KDL method. The execution times here are noticeably higher, registering at 53.67, 301.47, 830.02, and 2026.79 seconds for the same gradual increase in formal concepts. It's clear that as the number of formal concepts expands, the KDL method's execution time climbs drastically, indicating an intensifying computational requirement and limited scalability when tasked with larger datasets.

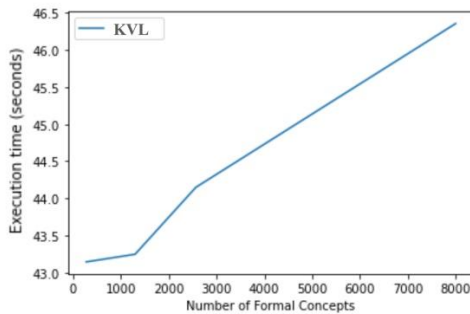


Fig. 10. Scalability of KVL method with increasing number of formal concepts.

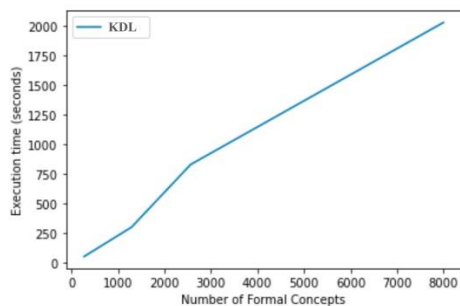


Fig. 11. Scalability of KDL method with increasing number of formal concepts.

This investigation rigorously assesses the K-means Dijkstra on Lattice (KDL) and K-means Vector on Lattice (KVL) algorithms, identifying a trade-off between clustering quality and computational efficiency. KDL excels in quality but is

resource-intensive, making it less scalable. Conversely, KVL is more scalable but may compromise on quality. The choice between the two hinges on task-specific needs: KDL is better for quality-focused tasks with sufficient resources, while KVL is ideal for tasks requiring scalability. Future research could aim to optimize each method's shortcomings, offering a more balanced clustering solution. These refinements would bring us closer to a unified, efficient, and high-quality clustering algorithm for handling categorical data.

## VII. CONCLUSION

In the exploration, the efficacy of a Dijkstra-based distance measure is assessed for conceptual clustering across multiple categorical datasets. This distance measure demonstrated a powerful capability in determining hierarchical relationships among categorical variables, even within complex and dense datasets. The evaluations, conducted across randomly generated formal contexts and real-world datasets, confirmed its robust performance, scalability, and reliability. However, a correlation between the average runtime and the number of concepts suggests potential efficiency enhancements.

The clustering tasks employed two methods: the K-means Dijkstra on Lattice (KDL) method, which uses Formal Concept Analysis (FCA) and the Dijkstra-based distance measure; and the K-means Vector on Lattice (KVL) method, which transforms categorical data into numerical vectors and applies standard k-means algorithms. The KDL method yielded high-quality clusters that accurately mirrored the inherent hierarchical relationships within categorical data. However, when handling larger numbers of clusters or formal concepts, scalability emerged as a challenge for this method. On the other hand, the KVL method demonstrated impressive scalability. Nevertheless, due to its conversion of data into numerical vectors, there's a risk of overlooking the hierarchical structure of the data, which could affect the clustering quality.

Future research has several promising pathways. The lattice structure in the KDL method could be simplified to boost scalability, and the KVL method could be further refined to better capture the structure of categorical data. Additionally, the exploration of alternate or complementary distance measures could be beneficial. A particularly intriguing direction for future research is integrating the Dijkstra-based distance measure into the k-means algorithm, which could significantly advance categorical data analysis. The study of the KDL and KVL methods has under-scored their respective strengths and limitations, illuminating potential areas for future research. These findings are instrumental to the ongoing development of categorical data analysis and refining data clustering methodologies. By investigating the complexities of concept lattices and streamlining the knowledge discovery process of FCA, The study offers a foundational understanding that serves as a basis for the development of more scalable and efficient solutions.

## ACKNOWLEDGMENT

The authors express sincere gratitude to the Department of Information Technology at the József Hatvany Doctoral School, University of Miskolc, Hungary, for the necessary support and resources for this study.



REFERENCES

- [1] V. Ganti, J. Gehrke, and R. Ramakrishnan, "CACTUS—clustering categorical data using summaries," in Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, 1999, pp. 73–83. <https://doi.org/10.1145/312129.312201>.
- [2] S. Guha, R. Rastogi, and K. Shim, "Rock: A robust clustering algorithm for categorical attributes," *Inf Syst*, vol. 25, no. 5, pp. 345–366, Jul. 2000, [https://doi.org/10.1016/S0306-4379\(00\)00022-3](https://doi.org/10.1016/S0306-4379(00)00022-3).
- [3] Z. Huang, "Extensions to the k-means algorithm for clustering large data sets with categorical values," *Data Min Knowl Discov*, vol. 2, no. 3, pp. 283–304, 1998. <https://doi.org/10.1023/A:1009769707641>.
- [4] Z. Huang, "Clustering large data sets with mixed numeric and categorical values," in Proceedings of the 1st pacific-asia conference on knowledge discovery and data mining, (PAKDD), Citeseer, 1997, pp. 21–34. <https://doi.org/10.4236/ojs.2017.72013>.
- [5] D. Ienco, R. G. Pensa, and R. Meo, "Context-based distance learning for categorical data clustering," in Advances in Intelligent Data Analysis VIII: 8th International Symposium on Intelligent Data Analysis, IDA 2009, Lyon, France, August 31-September 2, 2009. Proceedings 8, Springer, 2009, pp. 83–94. [https://doi.org/10.1007/978-3-642-03915-7\\_8](https://doi.org/10.1007/978-3-642-03915-7_8).
- [6] J. MacQueen, "Classification and analysis of multivariate observations," in 5th Berkeley Symp. Math. Statist. Probability, University of California Los Angeles LA USA, 1967, pp. 281–297. <https://doi.org/10.4236/ojpp.2015.56041>.
- [7] O. M. San, V.-N. Huynh, and Y. Nakamori, "An alternative extension of the k-means algorithm for clustering categorical data," *International journal of applied mathematics and computer science*, vol. 14, no. 2, pp. 241–247, 2004.
- [8] L. Chen and S. Wang, "Central clustering of categorical data with automated feature weighting," in Twenty-Third International Joint Conference on Artificial Intelligence, Citeseer, 2013.
- [9] M. Ng, M. Li, J. Huang, and Z. He, "On the Impact of Dissimilarity Measure in k-Modes Clustering Algorithm," *IEEE Trans Pattern Anal Mach Intell*, vol. 29, no. 3, pp. 503–507, Mar. 2007. <https://doi.org/10.1109/TPAMI.2007.53>.
- [10] R. Wille, "Restructuring lattices theory: an approach on hierarchies of concepts." Dordrecht, Holland: Springer, 1982.
- [11] R. Ganter and R. Wille, "Formal concept analysis: Mathematical foundations Springer-Verlag Berlin Germany," 1999. <https://doi.org/10.1007/978-3-642-59830-2>.
- [12] K. Sumangali and C. A. Kumar, "A comprehensive overview on the foundations of formal concept analysis," *Knowledge Management & E-Learning: An International Journal*, vol. 9, no. 4, pp. 512–538, 2017, <https://doi.org/10.34105/j.kmel.2017.09.032>.
- [13] M. Alwersh and L. Kovács, "Survey on attribute and concept reduction methods in formal concept analysis," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 30, no. 1, pp. 366–387, Apr. 2023, <https://doi.org/10.11591/ijeecs.v30.i1.pp366-387>.
- [14] T. Abiy, H. Pang, C. Williams, J. Khim, and E. Ross, "Dijkstra's shortest path algorithm," Retrieved from, 2016.
- [15] F. Mukhlif and A. Saif, "Comparative study on Bellman-Ford and Dijkstra algorithms," in *Int. Conf. Comm. Electric Comp. Net*, 2020.
- [16] R. Bellman, "On a routing problem," *Q Appl Math*, vol. 16, no. 1, pp. 87–90, 1958. <https://doi.org/10.1090/qam/102435>
- [17] R. W. Floyd, "Algorithm 97: shortest path," *Commun ACM*, vol. 5, no. 6, p. 345, 1962. <http://dx.doi.org/10.1145/367766.368168>.
- [18] M. Tropmann-Frick, "Analysis of the Shortest Path Method Application in Social Networks," 2023. <https://doi.org/10.3233/FAIA220500>.
- [19] T.-H. T. Nguyen and V.-N. Huynh, "A k-means-like algorithm for clustering categorical data using an information theoretic-based dissimilarity measure," in Foundations of Information and Knowledge Systems: 9th International Symposium, FoIKS 2016, Linz, Austria, March 7-11, 2016. Proceedings 9, Springer, 2016, pp. 115–130. [https://doi.org/10.1007/978-3-319-30024-5\\_7](https://doi.org/10.1007/978-3-319-30024-5_7).
- [20] Z. Huang and M. K. Ng, "A fuzzy k-modes algorithm for clustering categorical data," *IEEE transactions on Fuzzy Systems*, vol. 7, no. 4, pp. 446–452, 1999, <http://dx.doi.org/10.1109/91.784206>.
- [21] F. Cao, J. Liang, D. Li, L. Bai, and C. Dang, "A dissimilarity measure for the k-Modes clustering algorithm," *Knowl Based Syst*, vol. 26, pp. 120–127, 2012, doi: <https://doi.org/10.1016/j.knosys.2011.07.011>.
- [22] M. Li, S. Deng, L. Wang, S. Feng, and J. Fan, "Hierarchical clustering algorithm for categorical data using a probabilistic rough set model," *Knowl Based Syst*, vol. 65, pp. 60–71, 2014, doi: <https://doi.org/10.1016/j.knosys.2014.04.008>.
- [23] Bernhard Ganter, "Two basic algorithms in concept analysis. FB4-Preprint No 831, 1984." [https://doi.org/10.1007/978-3-642-11928-6\\_22](https://doi.org/10.1007/978-3-642-11928-6_22).
- [24] J. Baixeries, L. Szathmary, P. Valtchev, and R. Godin, "Yet a faster algorithm for building the Hasse diagram of a concept lattice," in Formal Concept Analysis: 7th International Conference, ICFCA 2009 Darmstadt, Germany, May 21-24, 2009 Proceedings 7, Springer, 2009, pp. 162–177. [https://doi.org/10.1007/978-3-642-01815-2\\_13](https://doi.org/10.1007/978-3-642-01815-2_13).
- [25] D. Schütt, "Abschätzungen für die Anzahl der Begriffe von Kontexten," Master's Thesis, TH Darmstadt, 1987.
- [26] L. Kovács, "Efficiency analysis of concept lattice construction algorithms," *Procedia Manuf*, vol. 22, pp. 11–18, 2018, <https://doi.org/10.1016/j.promfg.2018.03.003>.

# Intelligent Heart Disease Prediction System with Applications in Jordanian Hospitals

Mohammad Subhi Al-Batah<sup>1</sup>, Mowafaq Salem Alzboon<sup>2</sup>, Raed Alazaidah<sup>3</sup>

Faculty of Science and Information Technology, Jadara University, Irbid, Jordan<sup>1,2</sup>

Faculty of Information Technology-Department of CS, Zarqa University, Zarqa-Jordan<sup>3</sup>

**Abstract**—Heart disease is the leading cause of mortality worldwide. Early identification and prediction can play a crucial role in preventing and treating it. Based on patient data, machine learning techniques may be used to construct cardiac disease prediction models. This work aims to investigate the usage of machine learning models for heart disease prediction utilizing a publicly available dataset. The dataset contains patient information on clinical and demographic characteristics and the presence or absence of cardiac disease. Based on classification performance, many machine learning methods were tested and compared. The findings reveal that machine learning models can predict cardiac disease with accuracy and AUC values. Furthermore, the developed system is used to examine some Jordanian patients, and the predictions of the results are satisfactory. The study's findings might have far-reaching consequences for the early identification and prevention of heart disease, as well as for improving patient outcomes and lowering healthcare expenditures.

**Keywords**—Heart disease; machine learning; predictive models; classification; clinical data; predictions

## I. INTRODUCTION

Heart disease is a significant health concern that affects millions of people worldwide. Despite significant advancements in medical science, it remains a leading cause of death globally. Early detection and accurate heart disease prediction are crucial to prevent further complications and improve patient outcomes [1]. Machine learning techniques have shown promising results in several domains such as those in [2-6] and in predicting the occurrence of heart disease and other several diseases allowing for early intervention and better management [7]. Heart disease is a complex condition that involves various risk factors such as age, gender, heredity, smoking, hypertension, obesity, sedentary lifestyle, diabetes mellitus, metabolic syndrome, chronic renal failure, and stress. Identifying these risk factors and their interplay is critical in predicting the onset of heart disease. Traditional methods of heart disease prediction, such as clinical diagnosis, can be time-consuming, labor-intensive, and prone to errors. With the advent of machine learning techniques, accurate and efficient prediction of heart disease is now possible. The motivation behind this study is to develop a machine learning-based system that can effectively predict the occurrence of heart disease [2]. The system should be able to identify the most significant risk factors and their interplay, allowing for early intervention and better management. The study aims to improve upon existing methods of heart disease prediction and provide a reliable and efficient tool for healthcare

professionals. The primary research question of this study is: Can machine learning techniques effectively predict the occurrence of heart disease? Specifically, the study aims to answer the following sub-questions: What are the most significant risk factors for heart disease, and how do they interplay?

To answer this sub-question, the study will comprehensively analyze various risk factors associated with heart disease. For instance, the study may use data from electronic health records (EHRs) containing information on patients' demographics, medical history, lifestyle factors, and laboratory results [8]. The study may use statistical techniques such as logistic regression or correlation analysis to identify the most significant risk factors and their interplay [9]. For example, the study may find that smoking and hypertension are strongly associated with heart disease, and their co-occurrence increases the risk of heart disease even further. Next, which machine learning algorithms are most effective in predicting heart disease occurrence? To answer this sub-question, the study will evaluate the performance of various machine learning algorithms such as decision trees, random forests, support vector machines, and neural networks [10]. The study may use a dataset of patients with and without heart disease and train the models to predict the occurrence of heart disease. The study may use performance metrics such as accuracy, Precision, recall, and AUC to evaluate the models' performance [11].

The rest of the paper is organized as follows. Section II presents the relevant works with the subject under consideration. Additionally, in Section III, dataset description and analysis of the methodology are provided. Then, Section IV discusses the acquired research results. Finally, conclusions and future directions are outlined in Section V.

## II. LITERATURE REVIEW

To increase the accuracy of weak algorithms, an ensemble voting-based model combining many classifiers was presented. The proposed ensemble approach's power is appealing in increasing anemic classifier prognosis accuracy and establishing suitable performance in analyzing the risk of heart disease. An ensemble voting-based approach was used to achieve a remarkable gain in accuracy of 2.1 percent for anemic classifiers [12]. This study aims to create a model that can accurately forecast cardiovascular disorders to lessen the number of people who die from them. The suggested model was applied to a real-world dataset of 70,000 Kaggle instances and achieved the following accuracy: XGBoost: 86.87 percent

(with cross-validation), random forest: 87.05 percent (with cross-validation), multilayer perceptron: 87.28 percent (with cross-validation), and 86.94 percent (with cross-validation) (without cross-validation). AUC values for the suggested models are 0.94 for XGBoost, 0.95 for the random forest, and 0.95 for multilayer perceptron. According to the findings of this study, the multilayer perceptron with cross-validation surpassed all other algorithms in terms of accuracy, with the greatest accuracy of 87.28 percent [13].

Based on Machine Learning methods, this study provides an efficient and accurate solution for diagnosing cardiac disease. Several cutting-edge Machine Learning methods are used to classify a cardiovascular dataset. Machine Learning methods such as Random Forest, Nave Bayes, and SVM are used to introduce the prediction model. The prediction model is intended to provide improved performance with high accuracy [11]. This article examines the numerous machine learning algorithms utilized to accurately predict, diagnose, and treat various cardiac illnesses. The findings revealed that ANN had the highest average prediction accuracy (86.91 percent), whereas the C4.5 decision tree approach had the lowest average (74.0 percent). For automatic prediction, diagnosis, and treatment of heart disease, machine learning algorithms and techniques have been applied to several accessible heart disease datasets. The findings revealed that ANN had the highest average prediction accuracy (86.91 percent), whereas the C4.5 decision tree approach had the lowest average prediction accuracy (74.0 percent) [9]. Cardiovascular disease is a potentially fatal condition that has become more widespread in recent decades. Machine Learning tools and methodologies are employed to treat and diagnose this condition correctly. This study provides a survey of numerous models that accept such approaches and algorithms, as well as an analysis of their performance. Random Forest (RF), Decision Tree (DT), Naive Bayes, ensemble models, K-Nearest Neighbor (kNN), and Support Vector Machine are some common models (SVM) [14]. The heart is the human body's next main organ, and data analytics is utilized to anticipate the incidence of cardiac illnesses. To forecast the development of cardiac disorders, data mining and machine learning techniques such as Artificial Neural Network (ANN), Decision Tree, Fuzzy Logic, K-Nearest Neighbour (kNN), Nave Bayes, and Support Vector Machine (SVM) are utilized. This document includes an overview of known algorithms as well as a summary of previous work [14]. According to the World Health Organization, heart disease kills 33 percent of the world's population. To address this, a UCI Machine Learning Repository dataset was analyzed and predicted heart disease classes. The key characteristics of each assembling technique were retrieved, filtered from the dataset, fitted to a logistic regression classifier, feature scaling, and fitted using logistic regression. Mean Squared Error (MSE), Mean Absolute Error (MAE), R2 Score, Explained Variance Score (EVS), and Mean Squared Log Error (MSLE) were used to analyze performance (MSLE). The feature significance retrieved from the AdaBoost classifier was successful before adding feature scaling, with an MSE of 0.04, MAE of 0.07, R2 Score of 92 percent, EVS of 0.86, and MSLE of 0.16. The feature significance retrieved from the AdaBoost classifier was

successful after feature scaling, with an MSE of 0.09, MAE of 0.13, R2 Score of 91 percent, EVS of 0.93, and MSLE of 0.18 [15].

This research provides a machine learning framework that uses five algorithms to predict the likelihood of developing heart disease: Random Forest, Naive Bayes, Support Vector Machine, Hoeffding Decision Tree, and Logistic Model Tree (LMT). The Cleveland dataset is utilized for training and testing, and the findings demonstrate that Random Forest performs the best [16]. The essential information in this book is that machine learning and deep learning techniques are utilized to predict the range of risks connected with this project. A dataset was constructed by integrating previously accessible datasets and categorizing them into eleven groups. In diagnosing cardiovascular disorders, machine learning techniques outperformed deep learning, and the PCA methodology was used to determine the relative significance of each of the dataset's 11 fields. Random Forest Classifiers, Decision Tree Classifiers, and Naive Bayes algorithms surpassed other MI algorithms regarding accuracy and recalled. In undeveloped, developing, and even industrialized nations, heart disease is the leading cause of mortality. This disease's death rate can be reduced if detected early and accurately predicted. Machine Learning is critical in predicting and preserving key data regarding cardiac illnesses. This paper examines numerous research studies that use datasets to use machine learning in the prediction of cardiac ailments [17]. This will assist medical professionals in taking corrective action.

Cardiovascular diseases (CVDs) are the world's leading cause of sudden mortality today, and valid, accurate, and practical ways to identify them are required. Machine learning algorithms (MLAs) have been created and proven useful and efficient in forecasting CVD issues based on historical data. This thesis proposes a novel methodology that focuses on finding appropriate features by using MLA techniques such as Deep Learning, Random Forest, Generalized Linear Model, Naive Bayes, Logistic Regression, Decision Tree, Gradient Boosted trees, Support Vector Machine, Vote, and HRFLM, with higher accuracy levels of 75.8 percent, 85.1 percent, 82.9 percent, 87.4 percent, 85 percent, 86.1 percent, 78.3 percent, 86.1 percent, [18]. This research investigates the differences in performance of multiple machine learning models on chronic renal disease and cardiovascular disease datasets using Principal Component Analysis dimensionality reduction approaches. Logistic Regression, K Nearest Neighbor, Naive Bayes, Support Vector Machine, and Random Forest Model were utilized to assess the models' performance with and without PCA. The authors discovered that the kNN classifier and logistic regression were the best approaches for predicting renal and heart illness, with 100% accuracy in chronic kidney disease and 85% in heart disease [19]. This research investigates the differences in performance of multiple machine learning models utilizing Principal Component Analysis dimensionality reduction approaches on Chronic Kidney and Cardiovascular Disease datasets. Logistic Regression, K Nearest Neighbor, Nave Bayes, Support Vector Machine, and Random Forest Model were used to examine the performance of the models with and without PCA. The

scientists discovered that the kNN classifier and logistic regression were the best approaches for predicting renal and heart illness, with 100 percent accuracy in chronic kidney disease and 85 percent in heart disease, respectively [11].

Machine Learning is the study of how computers can learn to discover answers without being explicitly programmed. It is a subset of Artificial Intelligence that allows practitioners to identify illnesses more quickly and efficiently. Machine learning enables the creation of models that correlate various characteristics with an illness. However, good analytic tools are scarce for uncovering underlying correlations and patterns in data. Machine learning is becoming more prevalent in the diagnosis field due to the advancement of classification and recognition algorithms in disease categorization [20]. Data analysis is an essential element of healthcare since it allows for extracting hidden information and predicting disease. This research evaluates several machine learning approaches to determine the superiority of the Random forest algorithm in predicting heart disease [21].

### III. METHODOLOGY

The primary objective of this study is to identify an effective and predictive algorithm for the early detection of heart disease. To this end, we applied various machine learning models, including Logistic Regression (LR), Decision Tree, Neural Network, Naive Bayes (NB), the k-nearest neighbors (kNN), Support vector machines (SVM), AdaBoost (AB), Stochastic Gradient Descent (SGD), CN2 rule inducer, Constant and Random Forest (RF), to the Heart Disease Diagnostic dataset and evaluated the results. Fig. 1 outlines the planned architecture in detail.

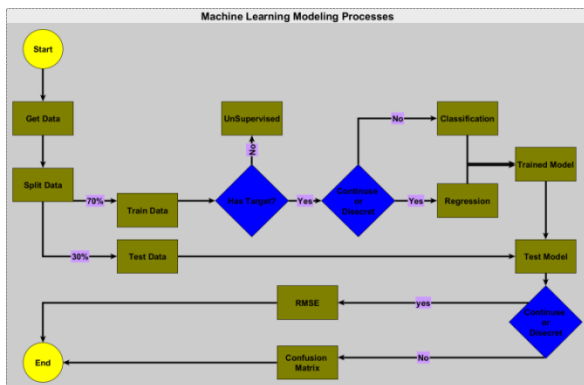


Fig. 1. Machine learning modeling processes flow diagram.

Our methodology begins with data collection, followed by preprocessing, which consists of four steps: data cleansing, attribute selection, target role setting, and feature extraction. Machine learning algorithms that can predict heart disease for a fresh set of measures are developed using the supplied data. To test the performance of an algorithm, we present the model with labeled new data. This is often accomplished by dividing the obtained labeled data into two sections using the Train test split function. Seventy-five percent of the data is referred to as the training data or training set and is utilized to construct our machine learning model. 25% of the data will be utilized to evaluate the model's performance, referred to as test data or test set. After evaluating the models, we compare the acquired

data to select the algorithm with the highest accuracy and discover the most predictive algorithm for heart disease detection in Fig. 2.

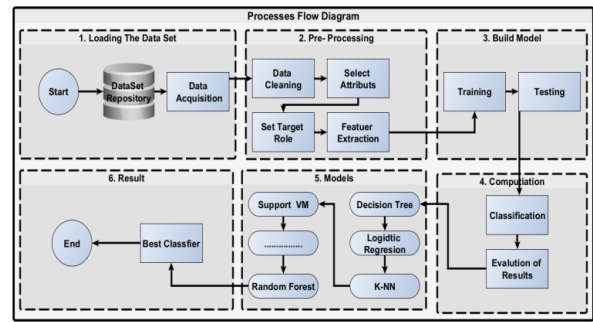


Fig. 2. Processes flow diagram.

#### A. Machine Learning Algorithms

The predictive analysis of machine learning algorithms is accomplished in our study. These are the machine learning algorithms used in our project.

Logistic regression is a statistical model used to estimate the probability of a binary outcome (such as yes/no) based on one or more independent variables. Logistic regression aims to determine the model that best represents the connection between the input features and the dependent variable. Logistic regression uses a logistic function to represent the relationship between the input features and the target variable, allowing for the calculation of the likelihood that the target variable would assume a specific value given the input information [22].

A decision tree model creates a tree-like structure by recursively dividing the data based on the input attributes. Each node in the tree indicates a determination based on a certain characteristic, while the leaves represent the final classification. Decision tree models are simple to read and depict, making them valuable for comprehending the model's decision-making process [23]. However, decision trees might be susceptible to overfitting if the tree is too complicated or if the data contains noise.

Neural Network: Neural networks are simulations of the structure and operation of the human brain. They are composed of layers of interconnected nodes (neurons) that identify data patterns and make predictions based on those patterns. Neural networks can learn intricate correlations between the input data and the target variable [24]. Neural networks can be challenging to comprehend and prone to overfitting if the model is too complicated or if there is insufficient data to train the model.

The Naive Bayes model calculates the probability of a certain class given the input features. It assumes that the characteristics are conditionally independent of one another, which simplifies the probability computation [25]. Naive Bayes is a simple and quick model that works well with tiny datasets, although, in some instances, it may not be as accurate as other models.

kNN is a non-parametric model that classifies new data points according to the class labels of the k nearest neighbours

in the training set. The value of  $k$  controls how many neighbours should be considered [26]. kNN is a simple and intuitive model that can perform well for low-dimensional data. Still, it can be computationally costly for high-dimensional data and may not perform well if the input is noisy or contains irrelevant features.

Support vector machines (SVM) are a model that locates a hyperplane that maximizes the difference between two classes in the data. Translating the input characteristics into a higher-dimensional space can handle both linearly and non-linearly separable data. SVM is a potent model that works well with high-dimensional data, but it is computationally intensive and may not scale well for huge datasets [27]. AdaBoost is a model that combines multiple weak classifiers to produce a powerful classifier. The final classification is based on the weighted combination of all the weak classifiers. AdaBoost is a potent model that can increase the accuracy of poor classifiers, but it is susceptible to noisy data and outliers [28]. SGD: Stochastic gradient descent (SGD) is a model that iteratively updates the model parameters using the gradient of the loss function for the model parameters. It modifies the parameters according to the gradients of the loss function to the parameters determined on a subset of the data (a mini batch). SGD is a quick and effective model that works well with large datasets but may require hyperparameter optimization for optimal performance [29]. CN2 is a model that learns decision rules from the input data. It employs a greedy search technique to determine the optimal collection of rules covering most situations while minimizing the number of rules. CN2 is a basic and interpretable model that works well with small datasets, but in some instances, it may not be as precise as other models [30]. Constant: The constant model is a model that always predicts the same class label for all instances. It is used as a baseline for comparison with other models and can be used to determine if a more complex model is necessary constant: The constant model predicts the same class label for all instances. It is used as a benchmark for comparing other models and to decide whether a more complex model is required [31]. Random Forest is a model that generates several decision trees based on random subsets of data and input attributes. The vote of most of all decision trees determines the highest classification. Random forest is a robust model that can handle high-dimensional data and noisy or irrelevant features. In addition, it is less susceptible to overfitting than decision trees. Random forest models can be challenging to interpret and computationally costly for large datasets [32]. Overall, each model has advantages and disadvantages, and the selection of a model depends on the nature of the problem being addressed and the data features [33]. It is essential to carefully pick and analyze the most suitable model for the work to get the highest potential performance.

### B. Dataset Acquisition

Data Collection and Preprocessing: Data collection entails getting the dataset from a trustworthy source, such as a medical research database or a publicly accessible dataset repository [34]. It is critical to ensure that the dataset is both relevant to the study issue and of acceptable quality. Data preparation is preparing a dataset for analysis by identifying

missing values, outliers, and inconsistencies. Missing values can be managed by imputation or removal, while outliers can be recognized and corrected via winsorization or trimming using statistical approaches or visualization techniques [35]. Inconsistencies can be fixed by looking for data input mistakes or integrating datasets. The models' performance on independent data can be evaluated by dividing the data into training and test sets. The training set trains models, while the test sets assess their performance [36]. It's critical to ensure the split is random and that the test set is representative of the entire dataset.

Extraction and Selection of Features: Categorical characteristics, such as gender or smoking status, have a restricted number of values. One-hot encoding is a method for representing category information as binary variables that may be fed into machine learning models [37]. Numeric characteristics are variables with continuous values, such as blood pressure or age. Scaling is a strategy to guarantee that the numeric characteristics have a consistent scale, preventing greater values from influencing the models [38]. A prominent scaling approach is standard scaling, which changes the data to have a mean of 0 and a standard deviation of 1.

Finding the most relevant characteristics for the classification task is known as feature selection. Correlation analysis may be used to find traits strongly connected to the target variable [39]. The feature importance ranking method may rank characteristics according to their relevance for the classification task.

Algorithms and Techniques for Machine Learning: Decision trees are a common machine-learning approach that builds a tree-like model of decisions and their potential outcomes [40]. Random forests are an ensemble approach that aggregates the forecasts of numerous decision trees.

Logistic regression is a machine learning approach that uses a logistic function to estimate the likelihood of a binary outcome. Support vector machines are machine learning algorithms that create a hyperplane to divide data into two classes. The K-nearest neighbours' method classifies data items based on the class of their  $k$  nearest neighbours [41]. Neural networks are a machine learning approach that uses a network of linked neurons to represent the connection between features and the goal variable.

Hyperparameter tuning entails picking the optimum values for the model parameters to maximise the model's performance on the data. This can be accomplished using grid search, random search, or Bayesian optimization techniques. Ensemble approaches integrate numerous models' predictions to increase their performance [42]. AdaBoost is an ensemble approach combining many weak models to get a stronger one. Bagging is an ensemble approach that uses bootstrap sampling to construct numerous models and then combines their predictions.

Metrics for Model Evaluation and Performance: Accurate is the proportion of correctly identified samples with the total number of samples. Precision is defined as the fraction of genuine positive samples among all positive samples. The

proportion of real positive samples out of the total number of positive samples is referred to as recall [43]. The harmonic mean of accuracy and recall is used to get the F1 score [44]. The area under the receiver operating characteristic curve (AUC-ROC score) evaluates the trade-off between genuine and false positive rates.

Cross-validation is a technique used to assess a model's capacity to generalize to new and previously unknown data [45]. It entails dividing the data into folds and testing the model on each fold while training it on the remaining folds [46]. Statistical tests may be used to assess the performance of several models and see if there are any significant differences [46]. T-tests or ANOVA (analysis of variance) can be employed to compare the means of two or more groups.

To illustrate the performance of the models, ROC curves and confusion matrices can be employed. ROC curves compare the true positive rate against the false positive rate at various threshold levels [47]. Confusion matrices display the model's true positive, true negative, false positive and false negative predictions.

### C. Experimental Environment

All machine learning algorithm tests mentioned in this research were conducted using Orange Data Mining version 3.35.0. The experimental machine was a Lenovo 20FES2FE0E with BIOS version N1GETA9W (1.88), an Intel Core i7-6600U CPU @ 2.60GHz (4 CPUs), and 8192MB of RAM. Windows 11 Pro 64-bit with build number 22621 was the operating system utilized for the experiment. The Heart Disease Diagnostic Datasets were used as the experimental dataset. The information supplied describes a dataset named "Heart Data Set." Here's a quick rundown of the dataset's characteristics:

The dataset has 1025 rows and 14 columns, which contain information about 1025 instances, each with 14 characteristics. The dataset has three categorical variables and ten numerical features, meaning that some features are qualitative (e.g., gender, kind of chest pain) and others are quantitative (e.g., age, resting blood pressure, serum cholesterol). The result variable is a categorical variable with two classes, indicating that the dataset is utilized for binary classification tasks. It isn't easy to interpret the dataset more thoroughly without further information about the target variable. It's worth mentioning that the dataset is popularly known as the Cleveland Heart Disease dataset and is frequently utilized for investigating predictive modeling jobs in the context of heart disease diagnosis.

## IV. RESULTS AND DISCUSSION

The information presented outlines a random sampling procedure used on a dataset of 1025 occurrences. Here's a quick rundown of the data:

The original dataset included 1025 occurrences, implying that it contained information about 1025 people. The random selection technique chose 75% of the data for inclusion, yielding a sample size of 769 occurrences. The sample is a subset of the original dataset that may be utilized for data analysis and modeling.

The remaining cases following the sampling procedure were not chosen for inclusion in the sample and were thus eliminated. In this example, the sample did not comprise 256 occurrences. It's worth noting that the rejected examples may include useful information that might impact the quality of any analysis or modeling conducted on the sample. As a result, it is critical to carefully analyze the sampling method employed and any potential biases imposed by the sampling process.

The material supplied provides a set of ten machine-learning models that may be utilized for various data analysis and modeling applications. Here's a quick rundown of each model:

Each model has advantages and disadvantages and may be better suited to tasks or datasets than others. Before picking the best model, it is critical to evaluate the problem at hand thoroughly, the nature of the data, and the limits of each model. Furthermore, it is critical to assess the model's performance using proper metrics and to interpret the findings with caution, considering any potential biases or restrictions caused by the data or modeling process as shown in Table I, Table II, and Table III.

TABLE I. DEMONSTRATES THE ACCURACY SCORES OF VARIOUS MACHINE LEARNING ALGORITHMS APPLIED TO A HEART DISEASE DETECTION DATASET

Model	AUC	CA	F1	Prec	Recall	MCC	Spec	LogLoss
RF	1	1	1	1	1	1	1	0.023
LR	0.938	0.874	0.873	0.876	0.874	0.749	0.869	0.321
Tree	0.871	0.867	0.867	0.876	0.867	0.744	0.874	4.581
SVM	0.999	0.973	0.973	0.973	0.973	0.945	0.972	0.069
AB	1	1	1	1	1	1	1	0
NN	0.983	0.94	0.94	0.94	0.94	0.88	0.939	0.176
kNN	1	1	1	1	1	1	1	0
NB	0.936	0.867	0.867	0.867	0.867	0.734	0.866	0.39
CN2	1	1	1	1	1	1	1	0.073
SGD	0.81	0.802	0.799	0.836	0.802	0.64	0.817	6.827

TABLE II. DEMONSTRATES THE ACCURACY SCORES OF VARIOUS MACHINE LEARNING ALGORITHMS APPLIED TO A HEART DISEASE DETECTION DATASET TARGET CLASS: 0

Model	AUC	CA	F1	Prec	Recall	MCC	Spec	LogLoss
RF	1	1	1	1	1	1	1	0.023
LR	0.938	0.874	0.861	0.901	0.825	0.749	0.918	0.321
Tree	0.871	0.867	0.87	0.812	0.937	0.744	0.804	4.581
SVM	0.999	0.973	0.971	0.973	0.97	0.945	0.975	0.069
AB	1	1	1	1	1	1	1	0
NN	0.983	0.94	0.936	0.947	0.926	0.88	0.953	0.176
kNN	1	1	1	1	1	1	1	0
NB	0.936	0.867	0.859	0.866	0.852	0.734	0.881	0.39
CN2	1	1	1	1	1	1	1	0.073
SGD	0.81	0.802	0.821	0.72	0.953	0.64	0.666	6.827

TABLE III. DEMONSTRATES THE ACCURACY SCORES OF VARIOUS MACHINE LEARNING ALGORITHMS APPLIED TO A HEART DISEASE DETECTION DATASET TARGET CLASS: 1

Model	AUC	CA	F1	Prec	Recall	MCC	Spec	LogLoss
RF	1	1	1	1	1	1	1	0.023
LR	0.938	0.874	0.884	0.853	0.918	0.749	0.825	0.321
Tree	0.871	0.867	0.864	0.934	0.804	0.744	0.937	4.581
SVM	0.999	0.973	0.974	0.973	0.975	0.945	0.97	0.069
AB	1	1	1	1	1	1	1	0
NN	0.983	0.94	0.944	0.934	0.953	0.88	0.926	0.176
kNN	1	1	1	1	1	1	1	0
NB	0.936	0.867	0.875	0.868	0.881	0.734	0.852	0.39
CN2	1	1	1	1	1	1	1	0.073
SGD	0.81	0.802	0.78	0.941	0.666	0.64	0.953	6.827

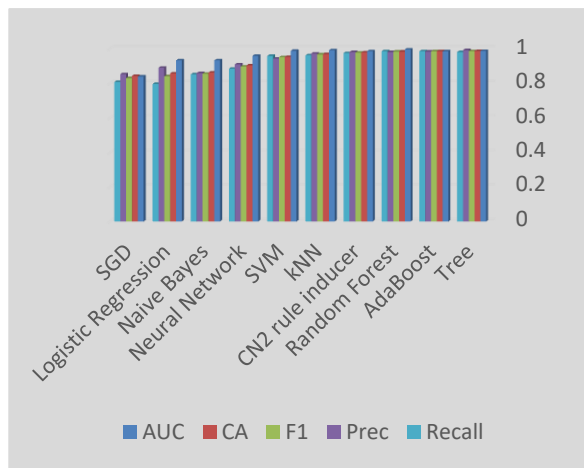


Fig. 3. Test and score, for the target class show: 0.

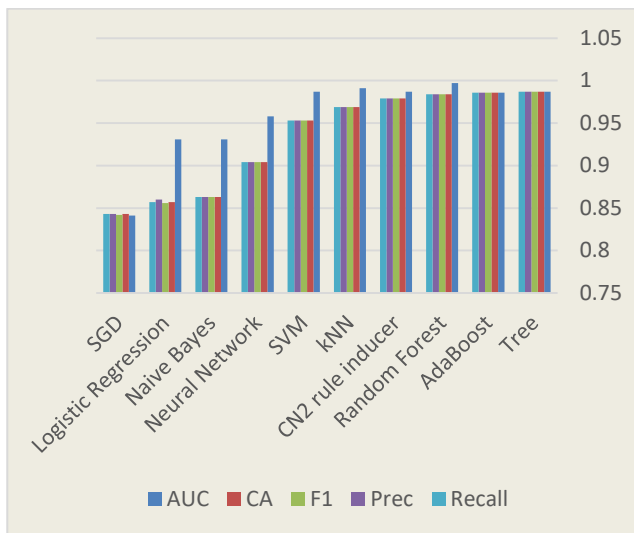


Fig. 4. Test and score, for the target class show average over classes.

The performance evaluation criteria for ten distinct machine learning models: interpretation and evaluation: The Random Forest model received excellent scores across all assessment measures, suggesting it did exceptionally well on the classification test. Fig. 4 shows test and score for the Target class Show: 0 whereas Fig. 3 shows the test and score

for the Target class show average over classes. However, it is important to note that the model may have been overfitting the data because it earned excellent scores on the training data. More testing on independent test data is required to demonstrate that the model generalizes successfully to new and previously unknown data. Logistic Regression: The Logistic Regression model performed well in most assessment measures, indicating that it fits the classification problem well. However, several criteria, such as Specificity and LogLoss, fell short, indicating that there may be space for development in these areas. Tree: The Tree model scored poorer than other models across all assessment measures, notably LogLoss. This suggests it may not be the most appropriate model for this classification problem. SVM: The SVM model received excellent scores in all assessment parameters, including AUC, F1, Precision, Recall, MCC, and Specificity. This suggests that it might be a good model for the classification problem. The AdaBoost model received perfect scores across all assessment measures, suggesting it did exceptionally well on the classification test. However, like with the Random Forest model, it must be evaluated on independent test data to ensure it generalizes effectively to new and unknown data. The Neural Network model scored moderate to high in most assessment measures but fell short in others, such as Specificity and LogLoss. This suggests that these areas may have space for improvement. The kNN model received perfect scores on all assessment criteria, suggesting it did exceptionally well on the classification test. However, like with the Random Forest and AdaBoost models, the model must be evaluated on independent test data to ensure it generalizes effectively to new and unknown data. Finally, the SGD model had the lowest scores across all assessment measures, suggesting that it may not be appropriate for this classification job, as shown in Fig. 5.

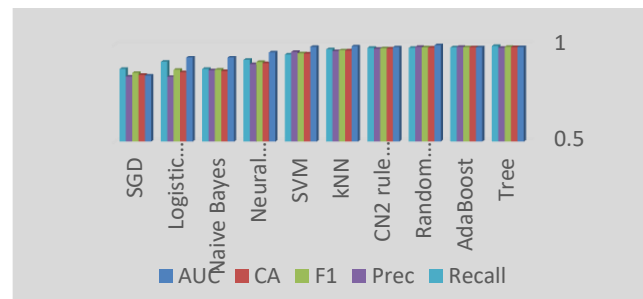


Fig. 5. Test and score, for the target class show: 1.

The information presented depicts the performance assessment metrics for 10 distinct machine learning models based on five evaluation metrics: AUC, CA, F1, Precision, and Recall. The models are assessed specifically on their ability to categorize data correctly and reliably as shown in Table IV.

According to the evaluation findings, the top-performing models are the Random Forest, AdaBoost, Tree, CN2 rule inducer, and kNN models, which received high scores across all assessment measures. The Random Forest model had the greatest AUC score, suggesting it has the best overall performance in rating the data. The Tree, AdaBoost, and CN2 rule inducer models all received perfect scores across all

assessment measures, suggesting they did exceptionally well on the classification test. The kNN model performed well across all assessment measures, indicating that it is good for the classification job.

TABLE IV. DEMONSTRATES THE ACCURACY SCORES OF VARIOUS MACHINE LEARNING ALGORITHMS APPLIED TO A HEART DISEASE DETECTION DATASET SHOW THE AVERAGE OVER CLASSES

Model	AUC	CA	F1	Prec	Recall
Tree	0.987	0.987	0.987	0.987	0.987
AB	0.986	0.986	0.986	0.986	0.986
RF	0.997	0.984	0.984	0.984	0.984
CN2	0.987	0.979	0.979	0.979	0.979
kNN	0.991	0.969	0.969	0.969	0.969
SVM	0.987	0.953	0.953	0.953	0.953
NN	0.958	0.904	0.904	0.904	0.904
NB	0.931	0.863	0.863	0.863	0.863
LR	0.931	0.857	0.856	0.86	0.857
SGD	0.841	0.843	0.842	0.843	0.843

The SVM model performed well in AUC, F1, Precision, and Recall, indicating that it might be a good model for the classification problem. However, it received a lower CA score, suggesting it may be less accurate in categorizing data than other models.

All assessment criteria gave the Neural Network model good ratings, indicating that it might be upgraded to increase its performance. Similarly, the Naive Bayes and Logistic Regression models performed well in most assessment criteria. Still, they performed poorly in others, such as Precision and Recall, showing space for development in these areas.

Finally, the SGD model had the lowest scores across all assessment measures, suggesting that it may not be appropriate for this classification job. The findings show that the Random Forest, AdaBoost, Tree, CN2 rule inducer, and kNN models are the best at this classification job. However, these models must be evaluated on independent test data to verify that they generalize effectively to new and unknown variables. Furthermore, it is critical to analyze the categorization task's unique goals and restrictions and choose the model that best satisfies those demands.

The information presented depicts the performance assessment metrics for 10 distinct machine learning models based on five evaluation metrics: AUC, CA, F1, Precision, and Recall. The models are assessed based on their ability to identify data appropriately and reliably. According to the assessment findings, the top-performing models are the Tree, Random Forest, AdaBoost, CN2 rule inducer, and kNN models, which received high scores across all evaluation measures. The Tree model had the greatest AUC score, suggesting it has the best overall performance in rating the data. All assessment metrics showed that the Random Forest, AdaBoost, and CN2 rule inducer models performed exceptionally well on the classification test. Except for F1, the kNN model had good scores in all assessment criteria, showing that it is viable for the classification job. The SVM

model scored well in AUC, Recall, and F1, indicating that it might be a good model for the classification problem. It did, however, have a lower Precision score, indicating that it may be less exact in correctly identifying data than some of the other models.

TABLE V. DEMONSTRATES THE ACCURACY SCORES OF VARIOUS MACHINE LEARNING ALGORITHMS APPLIED TO A HEART DISEASE DETECTION DATASET TARGET CLASS: 0

Model	AUC	CA	F1	Prec	Recall
Tree	0.987	0.987	0.986	0.992	0.981
AB	0.985	0.986	0.985	0.984	0.986
RF	0.996	0.984	0.984	0.981	0.986
CN2	0.986	0.979	0.978	0.981	0.975
kNN	0.991	0.969	0.967	0.972	0.962
SVM	0.988	0.953	0.951	0.943	0.959
NN	0.959	0.904	0.897	0.91	0.885
NB	0.933	0.863	0.856	0.859	0.852
LR	0.933	0.857	0.841	0.89	0.797
SGD	0.84	0.843	0.83	0.853	0.808

All assessment measures yielded moderate to low scores for the Neural Network, Naive Bayes, Logistic Regression, and SGD models, indicating that they may not be the optimal models for this classification problem. The assessment findings show that the top-performing models for this classification job are the Tree, Random Forest, AdaBoost, CN2 rule inducer, and kNN models. However, these models must be evaluated on independent test data to verify that they generalize effectively to new and unknown variables. Furthermore, it is critical to analyze the categorization task's unique goals and restrictions and choose the model that best satisfies those demands as shown in Table V.

Looking at the individual assessment measures, the Tree model had the greatest AUC score, suggesting it performs the best overall regarding data ranking. All assessment measures showed that the Random Forest, AdaBoost, CN2 rule inducer, and kNN models performed exceptionally well on the classification test. The CN2 rule inducer model had the greatest Precision score, meaning it is the most accurate at accurately identifying data. The Naive Bayes and Logistic Regression models have the highest Recall ratings, indicating they are the best at detecting true positives. The Neural Network model had the lowest scores across all assessment measures, implying that it is not the best model for this classification job. It's crucial to note that the performance assessment metrics reported here were calculated using a single random sample of the dataset, and the models' performance may change with various samples or on anonymous data. As a result, it is critical to carefully analyze the sampling method employed and any potential biases imposed by the sampling process. Furthermore, the quality of the data used to train and assess the models is important to their performance, and without knowing more about the dataset's origin, collection, and preprocessing, it's difficult to draw definitive conclusions or make suggestions based on this data alone.



The top-performing models for this classification job are the Tree, Random Forest, AdaBoost, CN2 rule inducer, and kNN models, with high scores across all assessment measures. The SVM model is also suitable, with excellent AUC, Recall, and F1 scores. The Naive Bayes and Logistic Regression models are the best at recognizing true positives, while the CN2 rule inducer model is the best at accurately categorizing data. The Neural Network model had the lowest scores across all assessment measures, implying that it is not the best model for this classification job. However, more testing on independent test data is required to demonstrate that these models generalize effectively to new and previously unexplored data. Furthermore, it is critical to analyze the categorization task's unique goals and restrictions and choose the model that best satisfies those demands.

TABLE VI. DEMONSTRATES THE ACCURACY SCORES OF VARIOUS MACHINE LEARNING ALGORITHMS APPLIED TO A HEART DISEASE DETECTION DATASET TARGET CLASS: 1

Model	AUC	CA	F1	Prec	Recall
Tree	0.987	0.987	0.988	0.983	0.993
AB	0.985	0.986	0.986	0.988	0.985
RF	0.996	0.984	0.985	0.988	0.983
CN2	0.986	0.979	0.98	0.978	0.983
kNN	0.991	0.969	0.97	0.966	0.975
SVM	0.988	0.953	0.955	0.962	0.948
NN	0.959	0.904	0.91	0.899	0.921
NB	0.933	0.863	0.871	0.867	0.874
LR	0.933	0.857	0.87	0.833	0.911
SGD	0.84	0.843	0.854	0.835	0.874

The information presented depicts the performance assessment metrics for 10 distinct machine learning models based on five evaluation metrics: AUC, CA, F1, Precision, and Recall. The models are assessed based on their ability to identify data appropriately and reliably. According to the assessment findings, the top-performing models are the Tree, Random Forest, AdaBoost, CN2 rule inducer, and kNN models, which received high scores across all evaluation measures. The Tree model had the greatest AUC score, suggesting it has the best overall performance in rating the data. All assessment metrics showed that the Random Forest, AdaBoost, and CN2 rule inducer models performed exceptionally well on the classification test. Except for Precision, the kNN model received good scores in all assessment criteria, showing that it is viable for the classification job as shown in Table VI.

The SVM model received excellent AUC, Precision, and Recall scores, indicating that it might be a good model for the classification problem. However, it had a lower CA score, indicating that it may be less reliable in accurately categorizing data than other models. All assessment measures yielded moderate to low scores for the Neural Network, Naive Bayes, Logistic Regression, and SGD models, indicating that they may not be the optimal models for this classification problem. The assessment findings show that the top-performing models for this classification job are the Tree,

Random Forest, AdaBoost, CN2 rule inducer, and kNN models. However, these models must be evaluated on independent test data to verify that they generalize effectively to new and unknown variables. Furthermore, it is critical to analyze the categorization task's unique goals and restrictions and choose the model that best satisfies those demands. Looking at the individual assessment measures, the Tree model had the greatest AUC score, suggesting it performs the best overall regarding data ranking. All assessment metrics showed that the Random Forest, AdaBoost, and CN2 rule inducer models performed exceptionally well on the classification test. The CN2 rule inducer model had the greatest Precision score, meaning it is the most accurate in accurately identifying data. The Neural Network model has the greatest Recall score, suggesting it is the best at detecting true positives. The Logistic Regression model has the greatest F1 score, suggesting a good mix of Precision and Recall.

The information supplied displays the classification results of 10 different machine-learning algorithms on a dataset of samples designated as malignant or benign as shown in Table VII. The findings are given as confusion matrices, which indicate the number of samples categorized as malignant or benign by the models and the samples' true labels. According to the assessment findings, the top-performing models include the Tree, Random Forest, AdaBoost, kNN, and CN2 models, which obtained high accuracy in accurately categorizing the data.

TABLE VII. SHOWS THE CONFUSION MATRIX OF SEVERAL MACHINE LEARNING ALGORITHMS APPLIED TO A HEART DISEASE DETECTION DATASET

Algorithms	Malignant	Benign	
Tree	358	7	Malignant
	3	401	Benign
Random Forest	360	5	Malignant
	7	397	Benign
Logistic Regression	291	74	Malignant
	36	368	Benign
SVM	350	15	Malignant
	21	383	Benign
AdaBoost	360	5	Malignant
	6	398	Benign
Neural Network	323	42	Malignant
	32	372	Benign
kNN	351	14	Malignant
	10	394	Benign
Naive Bayes	311	54	Malignant
	51	353	Benign
CN2	356	9	Malignant
	7	397	Benign
SGD	295	70	Malignant
	51	353	Benign

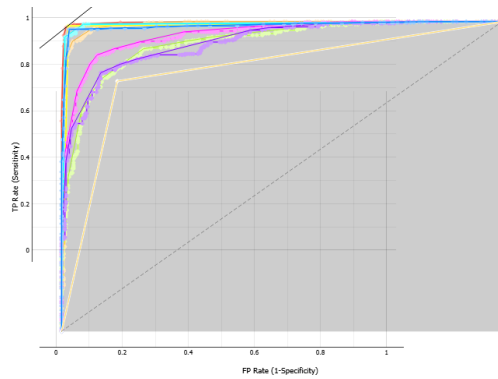


Fig. 6. The ROC curve to the total number of targets 0.

Fig. 6 depicts the ROC (Receiver Operating Characteristic) curve for target 0, a graphical depiction of a binary classification model's performance. The ROC curve is constructed by graphing the true positive rate (TPR) vs. the false positive rate (FPR) at different threshold levels.

The y-axis shows the TPR, also known as sensitivity or recall, which is the fraction of true positive cases accurately detected by the model. The x-axis depicts the FPR, the fraction of true negative cases the model mistakenly classifies as positive.

The ROC curve in the presented Fig. 6 is smooth and steep, indicating a well-performing model. The model's performance improves as the curve approaches the top-left corner of Fig. 6. This figure's curve is in the top-left corner, indicating that the model has a high TPR and a low FPR. In other words, the model can detect real positives while avoiding false positives.

The picture also includes the AUC (Area Under the Curve) score, which is a measure that highlights the overall performance of a binary classification model. The AUC score ranges from 0 to 1, with 1 representing perfect classification accuracy and 0.5 representing random guessing. In this scenario, the AUC value is near one, suggesting that the model performs well in categorizing the data with the goal 0.

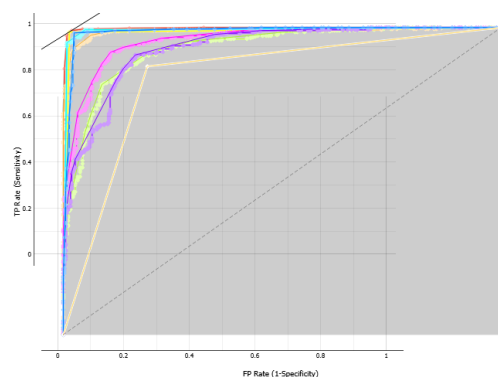


Fig. 7. The ROC curve to the total number of targets 1.

Fig. 7 depicts the ROC (Receiver Operating Characteristic) curve for target 1, a graphical depiction of a binary classification model's performance. The ROC curve is constructed by graphing the true positive rate (TPR) vs. the false positive rate (FPR) at different threshold levels.

## V. CONCLUSION AND FUTURE WORK

Finally, this work shows the promise of machine learning approaches for predicting heart disease using clinical and demographic data. The findings indicate that various machine-learning techniques may be utilized to create reliable predictive models for heart disease. The study's findings might have significant implications for the early identification and prevention of heart disease, as well as for improving patient outcomes and lowering healthcare expenditures. According to the findings of this study, machine learning models have the potential to be utilized as a tool for healthcare practitioners in the prediction of cardiac disease and the improvement of patient outcomes. These findings have major consequences for healthcare workers, patients, and healthcare systems. Early identification and prevention of cardiac disease can result in better patient outcomes and quality of life, lower healthcare costs, and more efficient resource allocation. Additional research and development are required to successfully integrate machine learning models in Jordanian hospitals. To ensure the generalizability and dependability of the predictive models, it is vital to evaluate the conclusions using separate datasets. Also required is research on the integration of machine learning models into existing healthcare systems and workflows. This includes addressing data quality, privacy, and interpretability issues, as well as designing user-friendly interfaces for healthcare professionals.

## REFERENCES

- [1] E. Elbasi, &A. I. Zreikat, A. I., "Heart Disease Classification for Early Diagnosis based on Adaptive Hoeffding Tree Algorithm in IoMT Data," *INTERNATIONAL ARAB JOURNAL OF INFORMATION TECHNOLOGY*, 20(1), 38-48. 2023.
- [2] 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.
- [3] R. Alazaidah, A. Al-Shaikh, M. R. AL-Mousa, H. Khafajah, G. Samara, M. Alzyoud, ... &S. Almatarnah, "Website Phishing Detection Using Machine Learning Techniques," 2024.
- [4] 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.
- [5] R. Alazaidah, &M. A. Almaiah, "Associative classification in multi-label classification: An investigative study," *Jordanian Journal of Computers and Information Technology*, 7(2). 2021.
- [6] R. Alazaidah, F. Thabtah, &Q. Al-Radaideh, "A multi-label classification approach based on correlations among labels," *International Journal of Advanced Computer Science and Applications*, 6(2), 52-59. 2015.
- [7] K. S. Jagadeesh and R. R., "Heart Disease Prediction Using Machine Learning," *Int. J. Res. Appl. Sci. Eng. Technol.*, 2022, doi: 10.22214/ijraset.2022.40918.
- [8] Deb, M. S. Akter Koli, S. B. Akter, and A. A. Chowdhury, "An Outcome Based Analysis on Heart Disease Prediction using Machine Learning Algorithms and Data Mining Approaches," in *2022 IEEE World AI IoT Congress, AIIoT 2022*, 2022, pp. 418-424. doi: 10.1109/AIIoT54504.2022.9817194.
- [9] L. Riyaz, M. A. Butt, M. Zaman, and O. Ayob, "Heart Disease Prediction Using Machine Learning Techniques: A Quantitative Review," in *Advances in Intelligent Systems and Computing*, 2022, pp. 81-94. doi: 10.1007/978-981-16-3071-2\_8.

- [10] D. S. Priyadarsini, "Heart Disease Prediction Using Machine Learning Algorithm," Vol. ISSUE-10, AUGUST 2019, Regul. ISSUE, 2019, doi: 10.35940/ijitee.j9340.0881019.
- [11] M. K. S. Ubale and D. P. N. Kalavadekar, "Effective Heart Disease Prediction Using Machine Learning Techniques," null, 2021, doi: null.
- [12] Javid, A. K. Z. Alsaedi, and R. Ghazali, "Enhanced accuracy of heart disease prediction using machine learning and recurrent neural networks ensemble majority voting method," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 3, pp. 540–551, 2020, doi: 10.14569/ijacsa.2020.0110369.
- [13] M. Bhatt, P. Patel, T. Ghetia, and P. L. Mazzeo, "Effective Heart Disease Prediction Using Machine Learning Techniques," *Algorithms*, vol. 16, no. 2, 2023, doi: 10.3390/a16020088.
- [14] M. Marimuthu, M. Abinaya, K. S., K. Madhankumar, and V. Pavithra, "A Review on Heart Disease Prediction using Machine Learning and Data Analytics Approach," *Int. J. Comput. Appl.*, vol. 181, no. 18, pp. 20–25, 2018, doi: 10.5120/ijca2018917863.
- [15] S. Basheer, R. M. Mathew, and M. S. Devi, "Ensembling Coalesce of Logistic Regression Classifier for Heart Disease Prediction using Machine Learning," *Int. J. Innov. Technol. Explor. Eng.*, doi: 10.35940/ijitee.I3473.1081219.
- [16] P. Motarwar, A. Duraphe, G. Suganya, and M. Premalatha, "Cognitive Approach for Heart Disease Prediction using Machine Learning," in *International Conference on Emerging Trends in Information Technology and Engineering, ic-ETITE 2020*, 2020. doi: 10.1109/ic-ETITE47903.2020.242.
- [17] Singh, D. Singh, and J. S. Samagh, "A Comprehensive Review Of Heart Disease Prediction Using Machine Learning," *J. Crit. Rev.*, vol. 7, no. 12, 2020, doi: 10.31838/jcr.07.12.63.
- [18] L. Chandrika and K. Madhavi, "A Hybrid Framework for Heart Disease Prediction Using Machine Learning Algorithms," in *E3S Web of Conferences*, 2021. doi: 10.1051/e3sconf/202130901043.
- [19] S. Mr, "A Detailed Analysis on Kidney and Heart Disease Prediction using Machine Learning," *J. Comput. Neurosci.*, 2021, doi: 10.53759/181x/jcns202101003.
- [20] Jayakrishnan, R. Visakh, and K. T. Ratheesh, "Computational Approach for Heart Disease Prediction using Machine Learning," *ICCIsc 2021 - 2021 Int. Conf. Commun. Control Inf. Sci. Proc.*, 2021, doi: 10.1109/ICCIsc52257.2021.9485014.
- [21] M. G, "Heart Disease Prediction Using Machine Learning Algorithms," *Biochem. Biophys. Res. Commun.*, 2020, doi: 10.21786/bbrc/13.11/6.
- [22] P. Singh and S. Agrawal, "Accuracy Prediction on Detection of Breast Cancer Using Machine Learning Classifiers," *Int. Conf. Comput. Intell. Commun. Networks*, 2022, doi: 10.1109/cicn56167.2022.10008345.
- [23] M. Abbas and H. Ghous, "Early Detection of Breast Cancer Tumors using Linear Discriminant Analysis Feature Selection with Different Machine Learning Classification Methods," *Comput. Sci. Eng. An Int. J.*, vol. 12, no. 1, pp. 171–186, 2022, doi: 10.5121/cseij.2022.12117.
- [24] M. C. Irmak, M. B. H. Tas, S. Turan, and A. Hasiloglu, "Comparative breast cancer detection with artificial neural networks and machine learning methods," in *SIU 2021 - 29th IEEE Conference on Signal Processing and Communications Applications, Proceedings*, 2021. doi: 10.1109/SIU53274.2021.9477991.
- [25] S. S. Olofintuyi, "Breast Cancer Detection With Machine Learning Approach," *Fudma J. Sci.*, 2023, doi: 10.33003/fjs-2023-0702-1392.
- [26] M. Mangukiya, "Breast Cancer Detection with Machine Learning," *Int. J. Res. Appl. Sci. Eng. Technol.*, 2022, doi: 10.22214/ijraset.2022.40204.
- [27] 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.
- [28] R. Alazaidah, G. Samara, S. Almatarnah, M. Hassan, M. Aljaidi, & H. Mansur, "Multi-Label Classification Based on Associations," *Applied Sciences*, 13(8), 5081. 2023.
- [29] M. Al-Batah, B. Zaqaibeh, 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.
- [30] Alshraiedeh, S. Hanna, & R. Alazaidah, "An approach to extend WSDL-based data types specification to enhance web services understandability," *International Journal of Advanced Computer Science and Applications*, 6(3), 88-98. 2015.
- [31] 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.
- [32] 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.
- [33] 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.
- [34] 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.
- [35] 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.
- [36] R. Alazaidah, F. K. Ahmad, & M. F. M. Mohsen, "A comparative analysis between the three main approaches that are being used to solve the problem of multi label classification," *International Journal of Soft Computing*, 12(4), 218-223. 2017.
- [37] 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>
- [38] K. Junoh, F. K. Ahmad, M. F. M. Mohsen, & R. Alazaidah, (2018, April). "Open research directions for multi label learning," in *2018 IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)* (pp. 125-128). IEEE. 2018.
- [39] 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>
- [40] 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.
- [41] 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>
- [42] M. S. Al-Batah, "Integrating the Principal Component Analysis with Partial Decision Tree in Microarray Gene Data," *IJCSNS International Journal of Computer Science and Network Security*, 19(3), 24-29, 2019.
- [43] M. S. Al-Batah, S. Mrayyen, and M. Alzaqebah, "Arabic Sentiment Classification using MLP Network Hybrid with Naive Bayes Algorithm," *Journal of Computer Science Science*, 14(8), 1104–1114, 2018.
- [44] R. Alazaidah, & F. K. Ahmad, "Trending challenges in multi label classification," *International Journal of Advanced Computer Science and Applications*, 7(10), 127-131. 2016.
- [45] M. Almomani, A. Momani, A. B. Abdelnabi, R. Zaqeba, and M. Al-Batah, "Predicting the corrosion rate of tempered medium carbon steel Using artificial neural network," *Protection of Metals and Physical Chemistry of Surfaces*, 58(2), 414-421, 2022.
- [46] S. Mrayyen, M. S. Al-Batah, and M. Alzaqebah, "Investigation of Naive Bayes Combined with Multilayered Perceptron for Arabic Sentiment Analysis and Opinion Mining," *International Journal of Mathematical Models and Methods in Applied Sciences*, 12, 2018.
- [47] M. S. Alzboon, M. S. Al-Batah, "Prostate Cancer Detection and Analysis using Advanced Machine Learning," (IJACSA) *International Journal of Advanced Computer Science and Applications*, 14(8), 388-396, 2023.

# A Novel Approach for Content-based Image Retrieval System using Logical AND and OR Operations

Ranjana Battur<sup>1</sup>, Jagadisha Narayana<sup>2</sup>

Dept. of Computer Science and Engineering-Karnataka Law Society's Gogte Institute of Technology,  
Visvesvaraya Technological University, Belagavi, India<sup>1</sup>

Dept. of Information Science and Engineering, Canara Engineering College, Mangalore Karnataka, India<sup>2</sup>

**Abstract**—This paper proposes an innovative ensemble learning framework for classifying medical images using Support Vector Machine (SVM) and Fuzzy Logic classifiers. The proposed approach utilizes logical AND and OR operations to combine the predictions from the two classifiers, aiming to capitalize on the strengths of each. The SVM and Fuzzy Logic classifiers were independently trained on a comprehensive database of medical images comprising various types of X-ray images. The logical OR operation was then used to create an ensemble classifier that outputs a positive classification if either of the individual classifiers does so. On the other hand, the logical AND operation was used to construct an ensemble classifier that outputs a positive classification only if both individual classifiers do so. The proposed method aims to increase sensitivity and precision by capturing as many positive instances as possible, thereby reducing false positives. The scope of the proposed work is validated in terms of overall time complexity and retrieval accuracy. The simulation outcome shows promising result with 98.36 accuracy score and 1.8 seconds to retrieve all the images in query database.

**Keywords**—Medical images; support vector machine; fuzzy logic; X-ray images; time complexity

## I. INTRODUCTION

The digital age has led to an explosion of visual content, from personal photos to professional databases. This has created a challenge for users who want to find specific images based on their content. Content-based image retrieval (CBIR) systems are a valuable tool for addressing this challenge [1]. CBIR systems work by extracting visual features from images, such as color, texture, and shape. These features are then used to compare images and find the best matches [2]. This is in contrast to traditional methods of image retrieval, which rely on textual metadata such as tags and keywords. CBIR is especially useful in domains where manual tagging is impractical or where the sheer volume of images makes keyword-based searching insufficient [3]. For example, CBIR can be used to find images of medical conditions, products, or people, even if they are not tagged with the relevant keywords [4]. The uses of CBIR are manifold. For example, in medical imaging, CBIR can help diagnose disease by comparing patient scans to an annotated image database [5]. In the domain of digital libraries, museums, and archival systems, CBIR facilitates the classification, indexing, and retrieval of visual content. Furthermore, e-commerce platforms have the potential to enable customers to search for products using images [6]. The urgency of efficient CBIR systems stems from our

increasing reliance on digital visuals [7]. As the volume of digital images continues to grow, the need for intelligent, content-based systems becomes even greater. They not only make the retrieval process more efficient but also open up avenues for more nuanced, context-aware searches, which are often lacking in traditional methods [8].

Despite its potential, CBIR systems face several challenges. One of the biggest challenges is the semantic gap. This is the difference between the low-level visual features extracted by CBIR systems and the high-level semantics understood by users [9-10]. For example, a CBIR system might consider two images to be identical based on their visual features, even if they contain different objects or are taken in different conditions. This can lead to mismatches during retrieval, where the system returns images that are not relevant to the user's query. Another challenge for CBIR systems is the heterogeneity of images. This refers to the fact that images can vary in terms of their angle, scale, lighting conditions, and other factors. This can make it difficult for CBIR systems to accurately match images, even if they contain the same objects [11]. Additionally, images often contain multiple objects, and it can be difficult for CBIR systems to identify the subject of interest [12]. This is especially true for images that are cluttered or have a lot of background noise.

However, even with these advances, there are still challenges to overcome. One challenge is ensuring that CBIR systems are both accurate and time-efficient. This is especially important in specialized domains such as medical imaging, where speed is often critical. Additionally, there is still a gap between machine-extracted features and human interpretation. This can lead to mismatches during retrieval, where the system returns images that are not relevant to the user's query. One possible solution to these challenges is to use an efficient set of classifiers. Our work aims to solve the challenges of CBIR in medical imaging by using an innovative combination of SVM and fuzzy logic. The study believes that this approach can bridge the semantic gap and improve the accuracy and efficiency of CBIR systems in this domain.

This paper introduces a novel approach to CBIR by amalgamating SVM and Fuzzy Logic, harnessing their combined strengths for enhanced image retrieval in the medical domain. The ensemble method, pivoting on logical operations, promises to address the challenges outlined above. The following are the key contributions:

- This paper presents a novel approach to CBIR integrating SVM and fuzzy logic.
- The proposed method aims to bridge the semantic gap that exists between the user's low-level visual features and high-level semantic understanding.
- This article introduces logical AND and OR operations as integration methods within a system. This allows switching between sensitivity and precision based on specific requirements, providing flexibility and adaptability for image retrieval tasks.
- Conduct in-depth analysis of time complexity to set benchmarks for real-time image capture. This contribution is significant, especially in critical areas such as medical imaging where time efficiency is critical.

Following this introduction, the remainder of this paper is organized in a structured manner as follows: Section II offers a brief survey of the existing work, presenting an overview of various methodologies and technologies currently in practice; In Section III, architectural design of the proposed system is discussed along with elucidating the processes of feature database creation, query image formulation, feature extraction, and the intricacies of the retrieval system; Section IV, discusses the system implementation, shedding light on the integration of SVM and Fuzzy Logic in CBIR, and the amalgamation of Ensemble SVM and Fuzzy System using Logical Operations; Section V, present a comprehensive evaluation of our system against a medical imaging database, highlighting the improved precision and recall rates. Finally, Section VI concludes the paper, summarizing the key findings, contributions, and implications of our work. It also outlines the potential avenues for future research, discussing the scalability, adaptability, and possible enhancements to our CBIR system.

## II. RELATED WORK

CBIR became known in the early 1990s and gained a lot of attention in research. Over the past decades, different feature extraction techniques have been employed based on appearances such as color, boundary contour, texture, and spatial layout. The work carried out by Younus et al. [13] devised a CBIR system considering color and texture-based feature descriptors and employed a joint approach of k-means and particle swarm optimization (PSO) for image retrieval. The presented scheme is validated on the WANG dataset. Based on the experimental outcome, it is identified that the precision for most classes is improved, but overlooked shape in similarity computations. The study conducted by Sajjad et al. [14] introduced a CBIR system customized to remain invariant to alterations in texture rotation and color features. For color feature extraction, they opted for the HSV color space, and a color histogram was employed for quantization. Notably, to circumvent the effects of illumination changes, only the Hue and Saturation channels were considered. To address rotation variance in textures, they utilized Local Binary Patterns (LBP). The effectiveness of the presented approach is justified based on three datasets namely, Zurich Building, Corel 1 K, and Corel 10 K.

Ashraf et al. [15], combined texture and color in a CBIR approach using HSV color moments, DWT, and Gabor wavelet. Their 1x250 dimensional feature vector improved accuracy but at the cost of slowed searches. The validation of the presented method is done against Corel and GHIM-10 K datasets. Based on the overall outcome analysis it has been identified that the presented method has achieved high precision, but also missing some texture and spatial details. The work presented by Alsmadi et al. [16] addresses the limitations of current CBIR systems by presenting a different scheme that extracts and stores comprehensive color, shape, and texture features as vectors. Techniques include RGB color paired with neutrosophic clustering, YCbCr color with wavelet transform, and gray-level matrices for texture. Utilizing a metaheuristic algorithm for similarity evaluation, the presented model demonstrated superior precision and recall against existing systems. The work of Choe et al. [17] presented an approach leveraging deep learning to retrieve similar chest CT images, aiming to assist examiners in the ILD (interstitial lung disease) diagnosis.

The work carried out by Garg and Dhiman [18] introduces a content-based image retrieval technique that combines discrete wavelet transformation with a rotationally invariant texture descriptor, enhancing feature extraction and reduction. By leveraging magnitude data and GLCM for texture classification, the approach, when tested on the CORAL dataset, demonstrated superior performance across classifiers like SVM, KNN, and decision trees in accuracy and other metrics. It has been observed that the existing CBIR methods falter in multi-class searches due to semantic similarities across image classes. Addressing this, the work of Khan et al. [19] introduces a hybrid CBIR method combining feature descriptors, a genetic algorithm, and an SVM classifier. By integrating color moments, various wavelets, and the L2 Norm for similarity measurement, the method effectively handles class imbalances. Tested on four datasets, it surpassed 25 CBIR techniques in retrieval performance. The authors in the study of Ma et al. [20] introduced a Privacy-preserving CBIR method for cloud-based multimedia, addressing current shortcomings in data encryption and feature extraction. Hybrid encryption is proposed alongside an enhanced DenseNet model for feature extraction from encrypted images. This ensures secure CBIR execution on cloud servers. Tested on two benchmarks, the method outperforms existing solutions by 1.9% and 10% in terms of accuracy, while also reducing computational costs and model parameters significantly.

Monowar et al. [21] introduced, a self-supervised image retrieval system using neural networks (NN), addressing the challenges of expensive or unfeasible data labeling. Trained on pairwise constraints, the model can function in self-supervised environments and with partially labeled datasets. It uses NN to extract and fuse image embeddings for retrieval. Banu et al. [22] explored CBIR using colour and texture features combined with high-level semantics. The proposed system outperforms traditional CBIR systems in speed and efficiency through an ontology model for content analysis when tested on a vast image database. The work of Rani in [23] presented a satellite image retrieval system for forest fire detection using hybrid feature extraction and a unique optimization algorithm

for feature selection; the method shows enhanced precision and recall, surpassing existing fire detection techniques.

Wang et al. [24] introduce a CBIR framework for skin diseases using multi-sourced information. This system fuses dermoscopic, clinical images, and meta-data, and with a graph-based community analysis, boasts a 0.836 precision on the EDRA and ISIC 2019 datasets. Arya and Vimina [25] presented the Local Neighborhood Gradient Pattern (LNGP) for CBIR, capturing local patterns in an 8-bit format. Tests on diverse datasets yield precision ranging from 40.66% to 86.12%, underlining its CBIR efficiency. Madhu and Kumar [26] unveil a hybrid algorithm for medical image feature extraction, merging techniques like DWT, PCA, and GLCM. Using various images, the approach is validated through K-means clustering, with an emphasis on enhancing classification precision via edge detection.

### III. PROPOSED SYSTEM

This section presents the proposed CBIR system and discusses the implementation procedure adopted to retrieve similar images from the image database. The prime aim of the research work reported in this paper is to evolve with an optimal approach of CBIR while achieving optimal decisions regarding content retrieval with higher precision and retrieval accuracy. In this regard, the study presents a unique and different approach from the existing work where the proposed study locally employs two different models to exploit their decision capability to get optimal results. The schematic depiction of the proposed CBIR system is shown in Fig. 1.

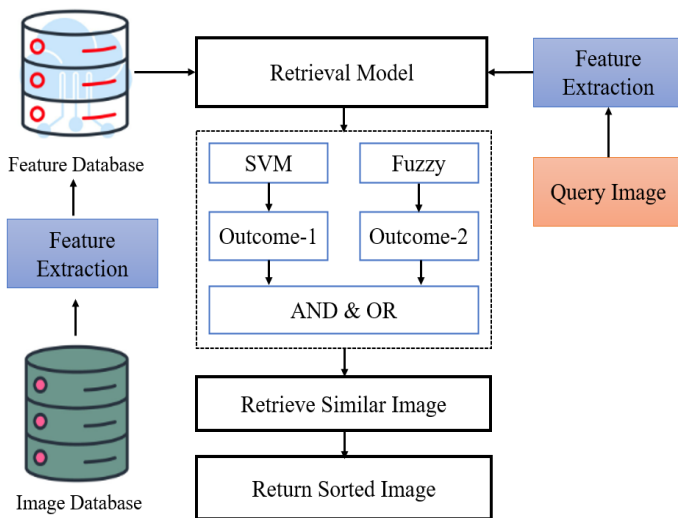


Fig. 1. Schematic architecture of the proposed CBIR system.

As shown in the above Fig. 1, the entire modeling of the proposed system can be divided into three core modules: i) construction of the feature database, ii) query image and its feature extraction, and iii) retrieval system itself. Each of these modules plays a pivotal role in ensuring the efficiency and efficacy of the entire process. Each of these modules plays a pivotal role in ensuring the efficiency and efficacy of the entire process. The system introduced in this paper emphasizes the need to optimize content retrieval, guaranteeing superior precision and accuracy.

#### A. Construction of Feature Database

This initial module plays a pivotal role in our CBIR approach. It involves meticulously processing the image database to extract and essential features. These features, which serve as unique identifiers for each image, form the foundation of our subsequent retrieval procedures. The system creates a feature-rich database that enables effective comparison and matching by capturing intrinsic characteristics like texture, colour, and shape. The database construction phase is twofold, customized to the distinct feature sets extracted for the SVM and Fuzzy Logic-based systems. For the SVM-based system, the feature extraction process revolves around three pivotal aspects:

- **CLD Coefficients:** Extracted to capture the spatial distribution of color in images, CLD coefficients provide insights into color layout variations. This feature is integral for distinguishing diverse medical image content.
- **Block Division:** By dividing images into smaller regions, block division enables localized analysis. This feature facilitates the identification of unique patterns and textures within specific image regions.
- **Edge Histogram:** Capturing the distribution of edges or gradients within images, the edge histogram feature highlights significant structural information. This attribute contributes to classifying images based on prominent edge distributions.

For the Fuzzy Logic-based system, the feature extraction process encompasses different elements:

- **Layer Segmentation:** The segmentation of images into distinct layers aids in identifying intricate structures within the medical images. These layers serve as crucial reference points for subsequent comparisons.
- **Region of Interest (ROI):** Extraction of ROIs isolates specific areas of medical interest. This feature guides the system towards recognizing critical regions with relevance to diagnosis.
- **Harris Corner and Fuzzy Corners:** Corner detection methodologies like Harris Corner and fuzzy corners enable the identification of distinct points in the image. These points, often representing unique features, enrich the feature database.

#### B. Query Image and Feature Extraction

The second module addresses the central challenge of CBIR—processing user queries effectively. Each query image presents a unique challenge and opportunity. This module employs advanced techniques to extract pertinent features from the query image. These extracted features are then organized into a structured representation that is well-suited for comparison with the features in our database.

#### C. Retrieval System

The third module is the heart of our CBIR system. Drawing on the extracted features and processed data from the previous

steps, this module executes the retrieval process. During the retrieval phase, the feature extraction process is applied to the query image. Subsequently, the SVM and Fuzzy Logic models independently process the query image, generating classification outputs based on their respective feature databases. In a harmonious ensemble, the outputs of both SVM and Fuzzy Logic classifiers are combined to derive the final retrieval decision. This amalgamation leverages the unique strengths of both models, enhancing the overall accuracy and robustness of the CBIR system. The diversity of features extracted from the two distinct databases ensures a comprehensive understanding of image content, enabling the system to provide more informed and precise retrieval outcomes.

#### IV. SYSTEM IMPLEMENTATION

This section discusses the implementation procedure adopted in the proposed system development.

##### A. SVM based CBIR

The SVM classifier used in our system is designed for CBIR which operates by extracting significant features from the images, specifically focusing on edge histogram, block division, and color layout descriptor (CLD) features. Block division allows the classifier to consider smaller regions of the image independently, providing local information that can further enhance the classification. CLD encapsulates the spatial distribution of color in the image, another crucial factor in differentiating between various images. It is one of the descriptors defined in the MPEG-7 standard, which is aimed at representing the spatial distribution of color in an image in a very compact form. CLD is highly effective for representing the spatial distribution of colors, even though the descriptor has a very small size. The implementation steps involved in the computation of CLD features are described as follows:

The algorithm begins by reading the input image, which is the target for color distribution analysis. To better represent color information, the image is converted from its original color space to YCbCr. This color space separation allows for a more effective analysis of color distribution. Further, the image is divided into non-overlapping blocks of size 8x8 pixels. If the image's dimensions are not evenly divisible by 8, padding is applied to ensure compatibility. Each block will be processed independently to capture color distribution characteristics. For every 8x8 block, the algorithm calculates the average color value. This is done in the YCbCr color space, where each pixel within the block contributes to the average.

---

##### Algorithm-1: CLD Feature Computation

---

###### Start:

1. Divide the image into non-overlapping 8x8 blocks
2. Get dimensions (h, w)
3. Blocks:  $N_{blocks} = \left\lfloor \frac{h}{8} \right\rfloor \times \left\lfloor \frac{w}{8} \right\rfloor$
4. If  $h \% 8 \neq 0$  or  $w \% 8 \neq 0$
5. do padding to get compatibility with the block size
6.  $\forall$  block  $B = 1:n$
7. compute the average color value
8.  $B_{i,j} \cdot C_{avg}^{i,j} = \frac{1}{8 \times 8} \sum_{y=0}^7 \sum_{x=0}^7 B_{i,j}(x,y)$

9. Apply 2D DCT to the 8x8 average color image

10. DCT coefficient for block  $B_{i,j}: D_{coeff}^{i,j}(u,v) = \alpha(u)\alpha(v) \sum_{y=0}^7 \sum_{x=0}^7 C_{avg}^{i,j}(x,y) \cos\left(\frac{(2x+1)u\pi}{16}\right) \cos\left(\frac{(2y+1)v\pi}{16}\right)$

11. The quantized coefficient for block

$$B_{i,j}: Q_{coeff}^{i,j}(u,v) = \text{round}\left(\frac{D_{coeff}^{i,j}(u,v)}{Q_{table}(u,v)}\right)$$

12. Arrange quantized DCT coefficients in 1D using zigzag scanning.

$$Z = [Q_{coeff}^{0,0}, Q_{coeff}^{0,1}, \dots, Q_{coeff}^{7,7}]$$

---

##### End

---

The computed average color values will represent the color essence of each block. The 2D DCT is employed on the average color values of each block. This mathematical transformation converts the spatial color information into frequency components, essential for capturing color distribution patterns. To reduce the amount of data while retaining crucial information, the DCT coefficients are quantized. Quantization involves dividing the coefficients by predefined quantization tables, which effectively reduces their precision. The quantized DCT coefficients are rearranged into a 1D array through zigzag scanning. This process converts the 2D array of coefficients into a compact, linear form, which is crucial for efficient storage and transmission of the descriptor. The resulting 1D array of zigzag-scanned quantized DCT coefficients forms the Color Layout Descriptor (CLD). This descriptor represents the image's color distribution in a highly compact yet informative manner. The computed CLD coefficients can now be used for various applications, such as image retrieval, content analysis, or classification, where the color distribution plays a pivotal role. The second feature namely edge histogram captures the local edge distribution in the image, which has proven to be a significant feature in distinguishing different types of medical images. The implementation steps involved in the computation of edge histogram features are subjected to multiple computing operations. The algorithm first loads the input image, which is then processed for edge histogram computation. To facilitate the process of edge detection, the algorithm transforms the image into grayscale via a weighted combination of its original color channels. This operation yields a single-channel representation where pixel values correspond to intensity levels. To identify edges within the grayscale image, an edge detection algorithm is engaged. This step may encompass techniques such as employing the Sobel or Canny algorithms, adept at highlighting zones of swift intensity fluctuations, often indicative of edges.

The outcome manifests as an edge map, delineating potential edge locations throughout the image. This edge map is then partitioned into discrete, non-overlapping blocks of a predefined magnitude, like 8x8 pixels. The division into smaller blocks fosters simplified analysis and a contextually confined comprehension of edge distribution within the image. For each of these blocks, the algorithm calculates an edge histogram, which quantifies the dispersion of edge orientations within the respective block. This procedure encompasses evaluating the intensity of each pixel's edge and allocating it into predefined bins associated with distinct edge orientations. The individual edge histograms derived from each block are

eventually amalgamated into a singular histogram, encompassing the collective distribution of edge orientations spanning the entire image. This unified histogram forms a succinct representation of how the edges are positioned within the image. The resulting composite histogram, the edge histogram, becomes a versatile asset applicable in diverse contexts. It effectively encapsulates the prevailing edge orientations present in the image and can be harnessed for tasks like image retrieval, object detection, or other undertakings that draw value from insights into the distribution patterns of edges.

---

**Algorithm-2:** Edge Histogram Computation

---

**Start:**

1. Convert the image to grayscale using a suitable color conversion formula
2. Grayscale intensity at pixel  $(x, y)$
3.  $I(x, y) = 0.29 \times R(x, y) + 0.58 \times G(x, y) + 0.11 \times B(x, y)$
4. Apply an edge detection algorithm (e.g., Sobel, Canny) to the grayscale image.
5.  $E(x, y)$  represents the edge map resulting from the algorithm
6. Divide the edge map into non-overlapping 8x8 blocks
7. Blocks:  $N_{blocks} = \left\lfloor \frac{h}{8} \right\rfloor \times \left\lfloor \frac{w}{8} \right\rfloor$
8. Edge histogram

$$H_{edge}^{i,j}(k): H_{edge}^{i,j}(k) = \sum_{y=0}^7 \sum_{x=0}^7 \delta(E_{i,j}(x, y), k)$$

9. Combine edge histograms of all blocks into a single

$$H_{combined}(k): H_{combined}(k) = \sum_{i=0}^{N_{blocks}} \sum_{j=0}^{N_{blocks}} H_{edge}^{i,j}(k)$$

---

**End**

---

Once the features are extracted, they are organized into a feature vector for each image. This vector becomes the input for the SVM that captures the image's unique characteristics. The extracted features are organized into a feature vector for each image. For example, consider there is  $m$  different features extracted for an image, then feature vector can be represented as follows:

$$x = x_1, x_2, x_3 \dots x_n \quad (1)$$

The combination of the extracted features into a single vector allows the SVM to effectively analyze the image data and make predictions based on these features. The SVM requires training to learn how to distinguish between different classes or categories. SVM learns to find a hyperplane that maximizes the margin between classes. The optimization problem involves finding the optimal  $\bar{w}$  and  $\bar{b}$  that define the hyperplane while considering misclassified points. SVM is a supervised learning model used for classification and regression tasks. The fundamental idea behind SVM is to find a hyperplane that best separates data points of different classes while maximizing the margin between these classes. This hyperplane serves as the decision boundary for classification, and the data points closest to the hyperplane are known as support vectors. Given a set of training data  $X = (x_1, y_1), (x_2, y_2) \dots (x_n, y_n)$ , where  $x_i$  represents the feature vector and  $y_i$  is the corresponding class label (either +1 or -1), the SVM aims to find a hyperplane  $w \cdot x + b =$ , that separates the two classes. The distance between the hyperplane

and the support vectors is given by the margin, which is proportional to  $1/||w||$ . The optimization problem for SVM can be formulated as follows:

$$\min: ||w||^2 + C \sum_{i=1}^n \xi_i \quad (2)$$

$$\text{subjected to: } y_i(w \cdot x + b) \geq 1 - \xi_i, s. t \xi_i \geq 0$$

where,  $C$  denotes a regularization parameter that balances the trade-off between maximizing the margin and minimizing the classification error, and  $\xi_i$  slack variables that allow for misclassified points or points within the margin. The goal is to minimize the value of  $||w||$ , while satisfying the classification constraints.

For image retrieval, a user provides a query image. The same feature extraction and vectorization process is applied to the query image. The SVM requires training to learn how to distinguish between different classes or categories. In the context of image retrieval, each image is assigned a label indicating its category. During training, the SVM constructs a decision boundary, known as a hyperplane that optimally separates the feature vectors of different classes. This hyperplane maximizes the margin between classes, effectively determining the most discriminative features for classification. Once the SVM is trained, it can be used for classification. For image retrieval, a user provides a query image that needs to be classified and compared against the images in the database. The query image undergoes the same feature extraction and vectorization process as the training images. The trained SVM then assesses the query image's feature vector and places it on the appropriate side of the decision boundary, classifying it into one of the predefined categories.

The SVM-based image retrieval process doesn't end with classification. Instead, it ranks the retrieved images based on their proximity to the decision boundary. Images that are closer to the decision boundary are considered more similar to the query image in terms of their feature characteristics. As a result, these images are ranked higher in the retrieval process. Finally, the retrieved images are presented to the user in the order of their ranking. Images that closely match the query image's features are displayed at the top of the list, providing the user with the most relevant results first.

**B. Fuzzy System-based CBIR**

The Fuzzy Logic-based classifier in our system focuses on edge detection for image retrieval. Edge detection is a fundamental tool in image processing and is critical in segmenting regions of interest in medical images. By employing Fuzzy Logic, the classifier can handle the inherent uncertainty and imprecision in edge detection, leading to a more robust image retrieval. The initial step involves extracting relevant features from the images. This could include processes such as median filtering, ROI extraction, and Harris corner detection. The outcome of this step is a collection of extracted features that serve as inputs to the subsequent fuzzy logic framework. Next, layer segmentation involves partitioning images into discrete layers. This process aids in the identification and isolation of intricate structures within medical images. Mathematically, the layer segmentation can be represented as dividing the image  $I$  into  $n$  distinct layers such



that:  $L_1, L_2, L_3, \dots, L_n$  where  $n$  denotes the number of layers; further, extracting Regions of Interest (ROIs) involves isolating specific areas within the images that hold medical relevance. Mathematically, an ROI can be defined as a subset of pixels within an image  $I$  defined by a set of coordinates  $(x, y)$  that denote the spatial bounds of the ROI. Corner detection techniques like Harris corners and fuzzy corners identify distinct points in the image. These points often correspond to unique features that contribute to the overall understanding of the image. Mathematically, Harris corner detection involves evaluating the corner response function  $R$  for each pixel in the image and identifying points with high corner responses. The implementation steps for computing Harris corner and fuzzy corners features are discussed in Algorithm 3.

The algorithm presented is a comprehensive approach for detecting both Harris corners and Fuzzy Corners within grayscale images. Starting with the Harris Corner Detection, the algorithm initially computes the gradients of the image using derivative filters, followed by the calculation of products of gradients and their smoothing through a Gaussian filter. The Harris response function is then computed for each pixel, reflecting the intensity of corner features. Non-maximum suppression is applied to identify potential corner locations, and a threshold is set to retain corners with significant Harris responses. However, what sets this algorithm apart is the incorporation of Fuzzy Corners. For each detected Harris corner, a degree of "cornerness" is calculated using a sigmoidal membership function (step-9). This degree of cornerness adds a layer of nuance to the corner detection, capturing the intensity of corners in a fuzzy manner. Linguistic variables are introduced to categorize corners based on their degree of cornerness, offering a more comprehensive understanding of corner characteristics. By combining traditional corner detection with fuzzy logic, the algorithm enhances the corner detection process. The inclusion of fuzzy degrees of corners brings a new dimension to corner characterization, allowing for a more nuanced representation of corner features within images. This approach is particularly valuable in scenarios where corners exhibit varying degrees of intensity, contributing to a richer understanding of image content. In essence, this algorithm presents a holistic framework for corner detection that embraces both crisp corner points and the fuzzy characterization of corner intensity, broadening the scope of corner analysis in image processing. The image retrieval process based on the provided steps involves a unique approach that combines fuzzy logic with the concepts of similarity measurement and defuzzification.

**Algorithm-3:** Harris Corner and Fuzzy Corners Detection

**Input:** Grayscale image  $I$  of size  $M \times N$ , Threshold  $T$  for corner intensity

**Output:** Sets of detected Harris corners and Fuzzy Corners

**Start:**

1. Convert the image  $I$  into gradients  $I_x$  and  $I_y$
2. Compute the products of gradients  $\forall$  pixel:  
 $I_x^2, I_y^2$  and  $I_{xy} = I_x \cdot I_y$
3. Apply a Gaussian filter to smooth the computed products:  
 $S_{I_x^2}, S_{I_y^2}$  and  $S_{I_{xy}}$
4. Compute the Harris response function for each pixel

$$R = \det(M) - k \times \text{trace}(M)^2$$

Where  $M$  is the matrix of second-order derivatives,  $k$  is an empirical constant (between 0.04 and 0.06),

$$\det(M) = S_{I_x^2} \cdot S_{I_y^2} - S_{I_{xy}}^2 \text{ and}$$

$$\text{trace}(M) = S_{I_x^2} + S_{I_y^2}$$

5. Apply non-maximum suppression to identify potential corner locations.
6. Set a threshold  $T$ , on  $R$  and retain corners with  $R$  above the threshold
7. Process: Fuzzy Corners
8.  $\forall C$
9. compute the degree of corners using a membership function:

$$\text{degreeofcornerness}(C) = \frac{1}{1 + e^{-\alpha \cdot R(C)}}$$

Where  $R(C)$  is the Harris response at corner  $C$ , and  $\alpha$  is a tuning parameter that controls the steepness of the membership curve.

10. Define linguistic variables low cornerness, moderate cornerness, and high cornerness using appropriate fuzzy sets.
11. Categorize each detected corner into one of the fuzzy sets based on its degree of corners
12. Return:  
Sets of detected Harris corners and their corresponding degrees of corners (fuzzy membership values)  
Categorized Fuzzy Corners based on their degree of corners

**End**

The adopted computing process offers a unique perspective on image similarity, enhancing the traditional methods used for content-based image retrieval. In the image retrieval, the degree of similarity between a query image and a database image is computed using fuzzy logic. This step capitalizes on the capability of fuzzy rule-based systems to handle imprecision and variability in data. Fuzzy rules are defined to establish relationships between the features of the query image and the database image, leading to the determination of similarity. The membership functions used in the proposed algorithm are defined based on the three factors viz. (i)  $\mu_Q(Q)$  Is the membership function for the query image's high corner intensity (HCI), (ii)  $\mu_D(D)$  is the membership function for the database image's HCI, and (iii)  $\mu_S(S)$  is the membership function for the degree of Similarity. Here, the variable  $Q$  represents the degree of HCI in the query image, similarly variable  $D$  represents the degree of HCI in the database image and variable  $S$  represents the degree of similarity between the query and database images. Using these variables and membership functions, the fuzzy rules are defined as follows:

$$\text{IF } Q = \text{High AND } D = \text{High, THEN } S = \text{HIGH} \quad (3)$$

$$\text{IF } Q = \text{Low AND } D = \text{Low, THEN } S = \text{Low} \quad (4)$$

These rules allow for a flexible interpretation of image similarity. If both the query image and the database image possess high corner intensities, their similarity is rated as high.

Conversely, if either the query or database image has low corner intensities, the similarity is considered low. However, the algorithm does not consider the logic for MEDIUM as it can lead to introduce redundancy, increase complexity, and may not add substantial value to the ranking and retrieval process. While the concept of a "Medium" category could theoretically represent a middle ground in corner intensities and hence, medium similarity, its practical implementation may not be as meaningful or effective given that the inherent numerical nature of intensities already provides a continuous scale of similarity. Therefore, the presented approach provides a nuanced way of assessing similarity based on multiple factors, leading to more contextually relevant image retrieval.

### C. Ensemble SVM and Fuzzy System using Logical Operation

The proposed ensemble learning approach ingeniously integrates the SVM and Fuzzy Logic classifiers by employing logical AND and OR operations. Given an input  $x$ , the outputs of the OR operation-based ensemble  $Y_{OR}(x)$  and the AND operation-based ensemble  $Y_{AND}(x)$  can be expressed as follows:

$$Y_{OR}(x) = SVM(x) \vee Fuzzy(x) \quad (5)$$

$$Y_{AND}(x) = SVM(x) \wedge Fuzzy(x) \quad (6)$$

where,  $SVM(x)$  represents the binary outputs of the SVM classifier and  $Fuzzy(x)$  refers to Fuzzy Logic classifiers, respectively. The symbols  $\vee$  and  $\wedge$  represent the logical OR and AND operations, respectively. In this new ensemble learning system, the OR-based ensemble predicts an instance to be positive if either the SVM or the Fuzzy Logic classifier predicts it to be positive. This arrangement increases the system's sensitivity by capturing more positive instances but also slightly increases the risk of false positives. Conversely, the AND-based ensemble predicts an instance to be positive only if both classifiers predict it to be positive. This increases the system's precision by reducing the risk of false positives but also slightly increases the risk of missing some true positives. This ensemble learning system aims to provide a more adaptable and robust solution for the challenging task of medical image classification by strategically leveraging the complementary strengths of both classifiers. In the OR operation, the ensemble essentially combines the outputs of both classifiers using majority voting. If the SVM or Fuzzy Logic classifier predicts a positive class, the ensemble also predicts a positive one. On the other hand, the AND operation involves constructing a meta-model that integrates the predictions of both classifiers. The ensemble only predicts a positive class if the SVM and Fuzzy Logic classifiers agree on the classification outcome. By intelligently fusing their decision-making capabilities through logical operations, the system ensures a more adaptable and resilient solution for the intricate task of medical image classification, addressing the need for sensitivity and precision in different clinical contexts. Algorithm 4 outlines a systematic process for combining SVM and Fuzzy Logic classifiers through ensemble learning. It involves training both classifiers using distinct features and then classifying test instances using both classifiers. The results from SVM and Fuzzy Logic classifiers are combined using logical operations to create two ensemble results. Depending on the outcomes of these ensemble results, instances are

classified as positive. This approach allows for harnessing the strengths of both classifiers, contributing to a more robust and flexible image classification system.

---

#### Algorithm-4: Ensemble Learning Using AND and OR

---

##### Inputs:

$D_{image}$  (Train Data attributed to Image Database)

$D_{query}$  (Test Data attributed to Query Database)

##### Outputs:

$\mathcal{R}_{OR}$  (results from the OR operation-based ensemble)

$\mathcal{R}_{AND}$  (results from the AND operation-based ensemble)

##### Start

1. Initialize Classifiers  
SVM classifiers:  $\mathcal{C}_{SVM}$   
Fuzzy logic classifier:  $\mathcal{C}_{FL}$
2. Classifier:  $\mathcal{C}_{SVM}$   
Extract feature  $\mathcal{F}_{SVM}$  including edge histogram, block division, and color layout  
Train  $\mathcal{C}_{SVM}$  using  $D_{image}$  and  $\mathcal{F}_{SVM}$
3. Classifier:  $\mathcal{C}_{FL}$   
Apply Fuzzy Logic-based edge detection on to obtain edge features  $\mathcal{F}_{Fuzzy}$   
Train Classifiers  $\mathcal{C}_{FL}$  using  $D_{image}$  and  $\mathcal{F}_{Fuzzy}$
4. Classify test instances  
For each instance  $x_i$  in  $D_{query}$ :  
Extract relevant features  $\mathcal{F}_i$  from instance  
Obtain SVM classification result:  
 $y_{SVM}(x_i) = \mathcal{C}_{SVM}(\mathcal{F}_i)$   
Perform Fuzzy Logic-based edge detection to extract edge features  $\mathcal{F}_{Fuzzy}^i$  from instance  
Obtain Fuzzy Logic classification result:  
 $y_{FL}(x_i) = \mathcal{C}_{FL}(\mathcal{F}_{Fuzzy}^i)$
5. Ensemble Classifications  
For each instance  $x_i$  in  $D_{query}$ :  
Compute AND ensemble:  
 $y_{AND}(x_i) = y_{SVM}(x_i) \wedge y_{FL}(x_i)$   
Compute OR ensemble:  
 $y_{OR}(x_i) = y_{SVM}(x_i) \vee y_{FL}(x_i)$
6. Return  
 $\mathcal{R}_{OR}$  consisting of all  $y_{OR}(x_i)$   
 $\mathcal{R}_{AND}$  consisting of all  $y_{AND}(x_i)$

##### End

---

## V. RESULT ANALYSIS

The proposed CBIR system was designed and developed using the Matlab and Python computing environments. The system was implemented on a Windows 10, core i7, system type 64-bit with 16GB RAM. The performance analysis of the proposed system is primarily carried out concerning retrieval accuracy and processing time with support of extensive discussion.

### A. Image Database Adopted

In this study a custom dataset comprising multi-modal images were created to evaluate the proposed CBIR system. The custom dataset was primarily based on the UNIFESP dataset, which contains 2,481 medical images. However, the study did not use all the images from this dataset. Instead, it used 1,500 images from UNIFESP and 858 images from other

sources. The experimental analysis was performed on a custom dataset of 1,920 images, which was divided into a training set (database) of 1,482 images and a test set (query database) of 438 images. The rationale behind this choice lies in the limited availability of comprehensive multi-modality datasets and as intended application of the proposed CBIR system is to facilitate analysis which should imitates real-world scenarios where users frequently search for diverse images. The database comprises a wide range of medical images, encompassing chest, hand, and ankle X-rays.

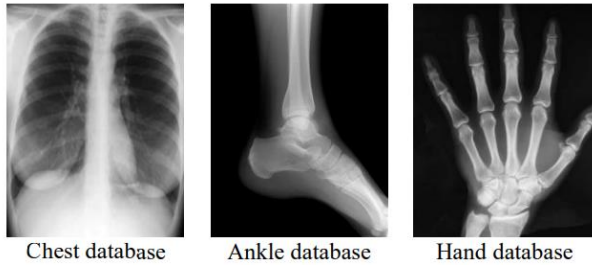


Fig. 2. Sample visualization of images in the database.

As shown in Fig. 2 above, the images are sourced from authoritative medical repositories and are characterized by their clinical relevance. Including medical images caters to the system's applicability in healthcare, enabling accurate retrieval of specific anatomical regions for diagnosis and reference.

### B. Visual Analysis of SVM Feature Descriptor

This section presents a visual analysis of the features descriptor obtained for the SVM-based CBIR system. The analysis provided for testing Ankle X-ray images from the query database.

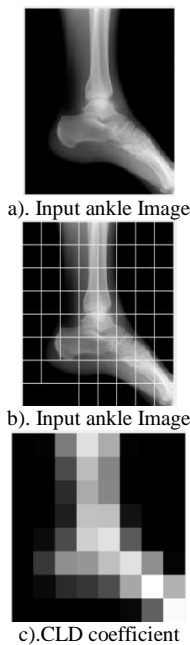


Fig. 3. Visual analysis of feature descriptor for SVM.

Fig. 3(a) shows that the initial image chosen for analysis is an Ankle X-ray image from the dataset. In Fig. 3(b) here, the same input Ankle X-ray image is displayed to provide a precise

reference for subsequent feature analyses. In Fig. 3(c), the Color Layout Descriptor (CLD) coefficients is exhibited, illustrating the spatial distribution of colours in the input Ankle X-ray image.

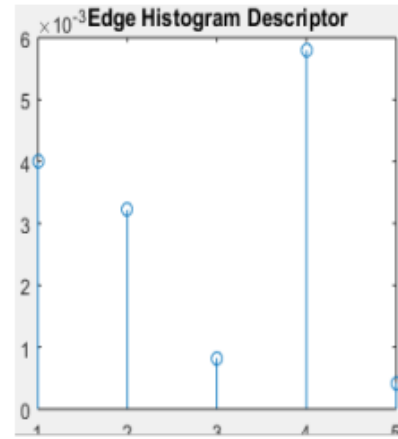


Fig. 4. Edge histogram descriptor

Fig. 4 present the edge histogram descriptor, which encapsulates the distribution of local edges in the image. This descriptor contributes to the system's ability to differentiate various image regions based on edge patterns. This visual analysis succinctly showcases the various feature descriptors generated by the SVM-based CBIR system. These descriptors are instrumental in enabling the system to capture distinct characteristics of the input image, facilitating effective content-based retrieval. This section gives readers insight into the transformative impact of feature extraction on the CBIR process.

### C. Visual Analysis of Fuzzy Feature Descriptor

This section presents a visual analysis of the features descriptor obtained for a fuzzy-based CBIR system. The analysis provided for testing Ankle X-ray images from the query database.

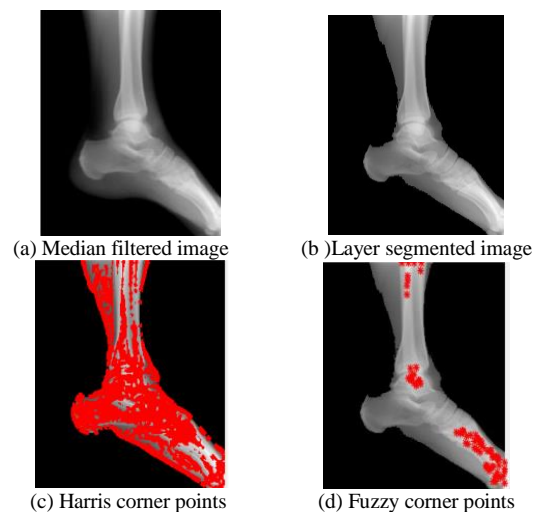


Fig. 5. Visual analysis of fuzzy feature descriptor.

In Fig. 5, a comprehensive visual analysis of the feature descriptors derived from the Fuzzy-based CBIR system, is

meticulously presented. Fig. 5(a) unveils the image after undergoing median filtering. This procedure smoothens the image, reducing noise while retaining essential structural information, rendering it an ideal input for subsequent analyses. Fig. 5(b) showcases the outcome of the layer segmentation process applied to the Ankle X-ray image. This segmentation facilitates the identification of distinct anatomical layers, facilitating precise feature extraction. The image depicted in Fig. 5(c) illustrates the Harris corner points, which are crucial for pinpointing distinctive features within the image. These points serve as significant landmarks in the feature extraction process. Fig. 5(d) brings forth the Fuzzy corner points, representing unique features identified within the image. These points contribute to the descriptor generation process, encapsulating salient attributes for retrieval. This visual analysis provides a profound understanding of the diverse feature descriptors synthesized by the Fuzzy-based CBIR system. These descriptors contribute to capturing intricate characteristics embedded within the input image. By navigating this section, readers better understand how feature extraction through Fuzzy Logic contributes to the CBIR process, enhancing the system's aptitude for content-based image retrieval.

#### D. Performance Analysis Concerning Retrieval Accuracy

The cornerstone of any CBIR system lies in its ability to retrieve relevant images based on user queries accurately. In the context of our proposed system, retrieval accuracy serves as a critical indicator of success. The assessment measures the proportion of retrieved images genuinely relevant to the query, thus quantifying the system's ability to discern and match image features. A high retrieval accuracy validates the efficacy of our approach in effectively capturing and exploiting distinctive features for content-based retrieval. The retrieval accuracy is calculated as a percentage, reflecting the proportion of correctly retrieved relevant images out of the total number of relevant images. This metric quantitatively assesses the CBIR system's ability to accurately identify and retrieve images that match the user's intent.

$$\text{Retrieval Accuracy} = \frac{\text{Number of Relevant Images Retrieved}}{\text{Total Number of Relevant Images}} \times 100$$

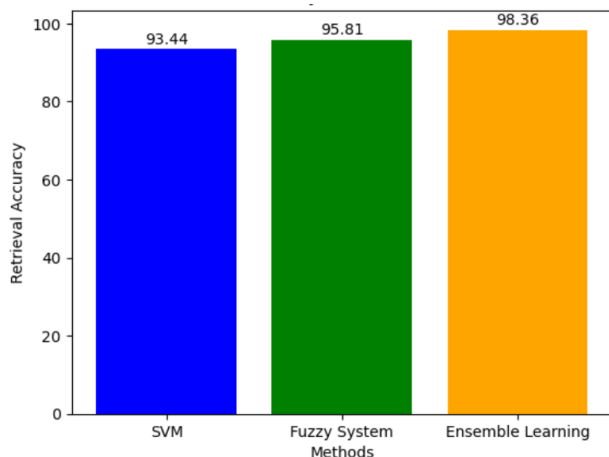


Fig. 6. Analysis of retrieval accuracy.

The results demonstrated in Fig. 6 clearly showcase the performance outcomes of each retrieval model, SVM, Fuzzy System, and Ensemble Learning. The SVM-based CBIR model demonstrates a retrieval accuracy of 93.44%. However, despite its high accuracy, SVM might encounter challenges in handling the intricacies of image data due to its linear nature. This limitation can lead to a slightly lower accuracy when compared to more adaptable models. The Fuzzy-based CBIR model exhibits an impressive retrieval accuracy of 95.35%. This higher accuracy can be attributed to the nature of Fuzzy Logic, which accommodates uncertainty and imprecision in image data. Fuzzy Logic is particularly well-suited for image segmentation and feature extraction, as it can handle varying degrees of membership. The Ensemble Learning-based CBIR model outshines the others with a remarkable retrieval accuracy of 98.36%. This higher accuracy can be attributed to the model's ability to harness the strengths of both SVM and Fuzzy Logic. The ensemble approach intelligently combines the decisions of these models using logical operations, striking a balance between precision and sensitivity. By doing so, it minimizes the weaknesses of individual models and leverages their combined potential. The SVM model provides a strong baseline, while the Fuzzy Logic model's adaptability enhances accuracy. The Ensemble Learning approach capitalizes on the synergies between SVM and Fuzzy Logic, leading to the highest accuracy due to its ability to adapt to different images and decision scenarios.

#### E. Performance Analysis Concerning Processing Time

While retrieval accuracy is paramount, it must be complemented by efficient processing time. The CBIR system's speed is integral, particularly in real-world scenarios where rapid responses are vital. The proposed system's processing time is evaluated as a measure of its computational efficiency. Lower processing times enhance user experience and render the system suitable for time-critical applications.

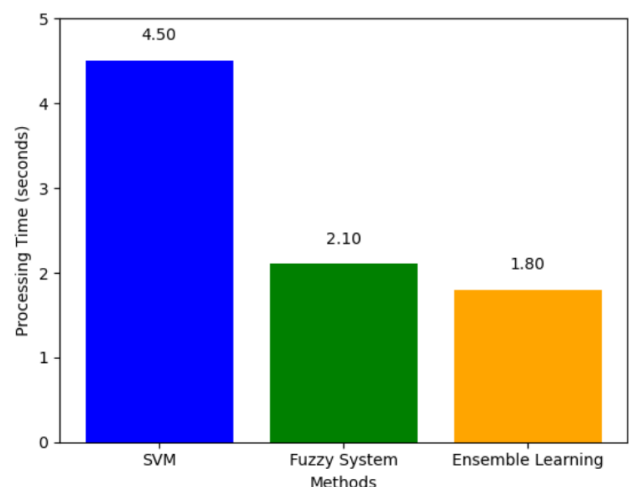


Fig. 7. Analysis of processing time.

The above Fig. 7 presents the performance analysis of the proposed system in terms of system response time for the different CBIR models, namely, SVM-CBIR, Fuzzy-CBIR, and Ensemble-CBIR. SVM-CBIR exhibits a longer processing time of 4.50 seconds on average. This can be due to the inherent complexity of the SVM's training process. Training of

SVMs involves iteratively optimizing the decision boundary that effectively separates the different classes. This process can lead to longer processing times in medical imaging scenarios, where datasets can be extensive and feature spaces complex. Fuzzy-CBIR, in contrast, exhibits a lower average processing time of 2.10 seconds. This efficiency is mapped to the nature of fuzzy logic operations, which handle membership degrees and linguistic variables. These operations enable quick calculations and quick decision-making. The processing efficiency of fuzzy-CBIR makes it particularly suitable for applications where quick retrieval with approximate similarity estimation is sufficient. The Ensemble-CBIR model balances accuracy and efficiency with an average processing time of 1.80 seconds. By combining the strengths of both SVMs and fuzzy logic classifiers through logical operations, the group adapts to the accuracy and sensitivity requirements. Although slightly higher than Fuzzy-CBIR due to Ensemble operation calculations, Ensemble-CBIR is more efficient than SVM-CBIR. This balance makes it an attractive solution for various medical image retrieval scenarios.

It is to be noted that the analysis of system processing time is also directly related to the time complexity analysis, which can be described as follows: The time complexity of the SVM-based CBIR model primarily revolves around two main phases: training and testing. The training phase involves extracting features from the training images and training the SVM model. If we denote the number of training images as 'n' and the number of features as m, the time complexity of feature extraction is typically  $O(n \times m)$ . The training of the SVM model will be  $O(n \times m^2)$ . Similar to the training phase, feature extractions and classification have a time complexity of  $O(m)$ . Several factors influence the time complexity of the Fuzzy Logic-based CBIR model. Edge detection and feature extraction involve processing each pixel of an image. If 'n' represents the number of pixels in an image, the time complexity for these steps is generally  $O(n)$ . Fuzzy Logic classification evaluates the fuzzy rules for each feature. The number of fuzzy rules and their complexity affect the time complexity, but it's typically  $O(1)$  for each feature. The time complexity for testing the Ensemble model depends on the testing time of both SVM and Fuzzy Logic classifiers and the time complexity of the ensemble operation (*OR* or *AND*). If we denote the time complexities of SVM and Fuzzy Logic testing as  $T_{svm}$  and  $T_{fuzzy}$ , respectively, and the time complexity of ensemble operation as  $T_{ensemble}$ , the overall time complexity can be approximated as  $O(T_{svm} + T_{fuzzy} + T_{ensemble})$ .

Hence, the time complexity analysis highlights that Fuzzy-CBIR generally has a lower time complexity due to its more effective feature extraction process. At the same time, SVM-CBIR and Ensemble-CBIR involve more complex feature extraction and classification steps.

## VI. CONCLUSION

This paper significantly contributes to content-based image retrieval (CBIR) systems. The authors explore a variety of methodologies, including support vector machines (SVMs), fuzzy logic, and ensemble learning, to develop an advanced CBIR system with potential applications beyond medical imaging. The proposed system's retrieval accuracy and

processing efficiency have been empirically validated, demonstrating its effectiveness in providing precise results promptly. The adaptable ensemble learning approach, which combines the strengths of established techniques, provides a balanced solution for diverse image retrieval scenarios. While this paper has made significant strides in the development of an effective and lightweight CBIR system, there remain several promising scopes for future research and improvement. In the future, the proposed work will be extended towards automated feature extraction process leveraging potential of the deep learning-based approaches to capture new and latent image characteristics, leading to even more precise retrieval results in dynamic and complex scenarios. In future, the study will evolve to incorporate multi-modal data combining medical images with textual patient records to offer a more comprehensive and context-aware CBIR system, especially in healthcare settings.

## REFERENCES

- [1] Latif et al., "Content-Based Image Retrieval and Feature Extraction: A Comprehensive Review," *Mathematical Problems in Engineering*, vol. 2019, pp. 1–21, Aug. 2019, doi: 10.1155/2019/9658350.
- [2] M. N. Abdullah et al., "Colour Features Extraction Techniques and Approaches for Content-Based Image Retrieval (CBIR) System," *Journal of Materials Science and Chemical Engineering*, vol. 09, no. 07, pp. 29–34, 2021, doi: 10.4236/msce.2021.97003.
- [3] Sarath Chandra Yenigalla, S. Rao, and Ngangbam Phalguni Singh, "Implementation of Content-Based Image Retrieval Using Artificial Neural Networks," *Engineering Proceedings*, Mar. 2023, doi: 10.3390/hmam2-14161.
- [4] S. Sikandar, R. Mahum, and A. Alsaman, "A Novel Hybrid Approach for a Content-Based Image Retrieval Using Feature Fusion," *Applied Sciences*, vol. 13, no. 7, p. 4581, Jan. 2023, doi: 10.3390/app13074581.
- [5] S. M. Anwar, M. Majid, A. Qayyum, M. Awais, M. Alnowami, and M. K. Khan, "Medical Image Analysis using Convolutional Neural Networks: A Review," *Journal of Medical Systems*, vol. 42, no. 11, Oct. 2018, doi: <https://doi.org/10.1007/s10916-018-1088-1>.
- [6] C. B. Akgül, D. L. Rubin, S. Napel, C. F. Beaulieu, H. Greenspan, and B. Acar, "Content-Based Image Retrieval in Radiology: Current Status and Future Directions," *Journal of Digital Imaging*, vol. 24, no. 2, pp. 208–222, Apr. 2010, doi: 10.1007/s10278-010-9290-9.
- [7] C. DeLorenzo, X. Papademetris, L. H. Staib, K. P. Vives, D. D. Spencer, and J. S. Duncan, "Image-Guided Intraoperative Cortical Deformation Recovery Using Game Theory: Application to Neocortical Epilepsy Surgery," *IEEE Transactions on Medical Imaging*, vol. 29, no. 2, pp. 322–338, Feb. 2010, doi: 10.1109/tmi.2009.2027993.
- [8] K. Juneja, A. Verma, S. Goel and S. Goel, "A Survey on Recent Image Indexing and Retrieval Techniques for Low-Level Feature Extraction in CBIR Systems," 2015 IEEE International Conference on Computational Intelligence & Communication Technology, Ghaziabad, India, 2015, pp. 67–72, doi: 10.1109/CICT.2015.92.
- [9] B. Ergen and M. Baykara, "Texture based feature extraction methods for content based medical image retrieval systems," *Bio-Medical Materials and Engineering*, vol. 24, no. 6, pp. 3055–3062, 2014, doi: 10.3233/bme-141127.
- [10] N. Borah and Udayan Baruah, "Feature Extraction Techniques for Shape-Based CBIR—A Survey," *Springer eBooks*, pp. 205–214, Dec. 2021, doi: 10.1007/978-981-16-4244-9\_16.
- [11] R. Battur and J. N., "CBIR System Development Using the Concepts of Image Feature Synthesis with Matching Parameters-A Survey," *Journal of Communication Engineering and its Innovations*, vol. 7, no. 1, Apr. 2021, doi: 10.46610/jocci.2021.v07i01.006.
- [12] B. Patel, K. Yadav, and D. Ghosh, "Current Trend and Methodologies of Content-Based Image Retrieval: Survey," *Algorithms for intelligent systems*, pp. 647–665, Jan. 2021, doi: 10.1007/978-981-15-6707-0\_64.

- [13] Z. S. Younus, D. Mohamad, T. Saba, H. M. Alkawaz, A. Rehman, M. Al-Rodhaan, and A. Al-Dhelaan, "Content-based image retrieval using PSO and k-means clustering algorithm," *Arabic Journal Geoscience*, vol. 8, no. 8, pp. 6211–6224, 2015.
- [14] M. Sajjad, A. Ullah, J. Ahmad, N. Abbas, S. Rho, and S. W. Baik, "Integrating salient colors with rotational invariant texture features for image representation in retrieval systems," *Multimedia Tools and Applications*, vol. 77, no. 4, pp. 4769–4789, Feb. 2018.
- [15] R. Ashraf, M. Ahmed, U. Ahmad, M. A. Habib, S. Jabbar, and K. Naseer, "MDCBIR-MF: Multimedia data for content-based image retrieval by using multiple features," *Multimedia Tools and Applications*, vol. 79, no. 13–14, pp. 8553–8579, Apr. 2020.
- [16] M. K. Alsmadi, "Content-Based Image Retrieval Using Color, Shape and Texture Descriptors and Features," *Arab J Sci Eng*, vol. 45, pp. 3317–3330, 2020.
- [17] J. Choe et al., "Content-based image retrieval by using deep learning for interstitial lung disease diagnosis with chest CT," *Radiology*, vol. 302, no. 1, pp. 187-197, 2022.
- [18] M. Garg and G. Dhiman, "A novel content-based image retrieval approach for classification using GLCM features and texture fused LBP variants," *Neural Comput & Applic*, vol. 33, pp. 1311–1328, 2021.
- [19] U. A. Khan, A. Javed, and R. Ashraf, "An effective hybrid framework for content based image retrieval (CBIR)," *Multimed Tools Appl*, vol. 80, pp. 26911–26937, 2021.
- [20] W. Ma et al., "A privacy-preserving content-based image retrieval method based on deep learning in cloud computing," *Expert Systems with Applications*, vol. 203, p. 117508, 2022.
- [21] M. M. Monowar et al., "AutoRet: A self-supervised spatial recurrent network for content-based image retrieval," *Sensors*, vol. 22, no. 6, p. 2188, 2022.
- [22] J. Faritha Banu et al., "Ontology Based Image Retrieval by Utilizing Model Annotations and Content," in *2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, Noida, India, 2022, pp. 300-305.
- [23] K. V. Rani, "Content based image retrieval using hybrid feature extraction and HWBMMBO feature selection method," *Multimed Tools Appl*, 2023.
- [24] Y. Wang et al., "Multi-channel content based image retrieval method for skin diseases using similarity network fusion and deep community analysis," *Biomedical Signal Processing and Control*, vol. 78, p. 103893, 2022.
- [25] R. Arya and E. R. Vimina, "Local neighborhood gradient pattern: A feature descriptor for content based image retrieval," *Journal of Intelligent & Fuzzy Systems*, vol. 43, no. 4, pp. 4477-4499, 2022.
- [26] R. Madhu and Kumar, "A hybrid feature extraction technique for content based medical image retrieval using segmentation and clustering techniques," *Multimed Tools Appl*, vol. 81, pp. 8871–8904, 2022.

# Imperative Role of Digital Twin in the Management of Hospitality Services

Ramnarayan<sup>1</sup>, Rajesh Singh<sup>2</sup>, Anita Gehlot<sup>3</sup>, Kapil Joshi<sup>4</sup>,  
Ashraf Osman Ibrahim<sup>5</sup>, Anas W. Abulfaraj<sup>6</sup>, Faisal Binzagr<sup>7</sup>, Salil Bharany<sup>8</sup>

Department of CSE-Uttaranchal Institute of Technology-Uttaranchal University, Dehradun-248007, India<sup>1, 2, 3, 4</sup>  
Creative Advanced Machine Intelligence Research Centre-Faculty of Computing and Informatics,  
Universiti Malaysia Sabah, 88400 Kota Kinabalu, Sabah, Malaysia<sup>5</sup>

Department of Information Systems, King Abdulaziz University, P.O. Box 344, Rabigh; 21911, Saudi Arabia<sup>6</sup>

Department of Computer Science, King Abdulaziz University, P.O. Box 344, Rabigh 21911, Saudi Arabia<sup>7</sup>

Department of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab 144402, India<sup>8</sup>

**Abstract**—Digital twin implementation enables more effective terms of evaluation and planning, and also effective utilization of resources with a flood of knowledge to improve the real-time services. The hospitality industry settings utilize digital twin technologies to introduce new ideas with sensor, actuators, AR/VR improve production, and improve customer services. Currently, the hospitality industry is focused to create a fast, virtual world space where customers can get a real world of hospitality. The technologically digital twin of a vast inn office can be implemented to create both discrete and continuous event recreations in order to precisely conceptualize the events that occur in distinct frameworks. Based on the above facts, the adoption of the digital twin in the hospitality industry has gained significant attention. With this motivation, the study aims to investigate the significance and application of the digital twins in the hospitality industry for establishing innovative and digital infrastructure. In addition to this, the study discusses different elements that are significant for the digital twin. Finally, the article summarizes and recommends vital recommendation in the adoption of digital twin in hospitality industry.

**Keywords**—Hospitality industry; digital twin; sensor and actuator; IoT; augment and virtual reality

## I. INTRODUCTION

The Sustainable Development Goals act as a road map for building a better and more sustainable future for all while also addressing urgent, serious global issues [1]. This article will examine some of the challenges that the hotel industry may encounter to assist it to contribute to the Sustainable Development Goals (SDGs) [2]. It will also provide some deeper, more thorough viewpoints on industrial sustainability. One of the primary hospitality-related SDGs of the 2030 Agenda of the United Nations is "sustainable tourism"[3]. Three SDGs specifically mention the hospitality industry: life below water (SDG 12), sustainable consumption and production (SDG 12), and sustainable economic growth (SDG 8) [4]. For many years, hospitality industry facing challenges in the consumer perception of the way to facilitate the services to improve the quality and reliability [5]. The hotel industry's assertiveness is influencing numerous substantial and small collaborators to leverage technologies to alleviate limitations [6]. The hospitality industry is undergoing a digital transformation that will result in a far more personalized,

customer-focused experience. Digital twin technologies are an indication of this development because they offer guests more flexibility over even the most modest components of their stays [7].

With the use of digital twins, restaurants, hotels and other hospitality businesses can get an advantage over rival businesses [8]. The majority of hotels employ connected devices nowadays to streamline customer requests and identify their unique characteristics in order to provide customized services. Examples of this include Marriot and other stakeholders' design of smart hotels and the use of concierge digital twins by various groups [9]. When used correctly, digital twin arrangements are capable of accurately recreating various resources, cycles, and frameworks in a virtual environment. This effectively makes the digital twin an option for huge accommodations were monitoring multiple cycles within a framework while executing innovative concepts is a frequent occurrence [10]. This diagram illustrates how the hospitality sector uses digital twins to portray digital pictures of all management.

This paper makes up a commitment to the top-of-the-line digital twins for putting together and coordinating operational frameworks for the hospitality industry. The primary goal of this study is to draw attention to the divergence between the theoretical and imaginary network of digital twins for assembly and storage and their practical application in considerations of theatrical illustration. The main contributions of the study are presented in the following:

- The most recent advancements in reenactment techniques that might establish the foundation for digital twins with advanced levels of information reconciliation, mechanization, and smart skills.
- This study uses a similar approach to analyze fictitious and specialized preparation for building highly devoted and powerful digital twins for the hospitality industry.
- Although digital twin technologies have shown significant advantages, their implementation still faces difficulties.

- In this paper, we are representing the future benefits of digital twins in this sector because new research opportunities have been made possible by the introduction of digital twin technology.

The study organized as follows: The background of the hospitality sector and digital twin covered in Section II, along with an overview of the technology in Section III, sensor and actuator presentations in Section IV, the Internet of Things in Section V, AR and VR in Section VI, and digital twin in the hospitality sector in Section VII.

#### A. Methodology

This section includes the methods used to complete the review on the digital twin in the hospitality industry. All the sections contain the analysis, data collection, criteria, and the searched strategy over the digitalization with digital twins. The concern of the review is to connect the digital twins and related technologies with the hospitality industry for future automation. The research question is: How digital twin technology can automate the hospitality industry? Based on this research question, we have collected research and review papers. The research paper was collected from various databases such as Scopus and web of science. The following parameters have been followed for the inclusion and exclusion of articles for analysis and they are: abstract of the paper but the full text of the study is not examined in the review. Algorithms and methodologies were used in the review but results are not used in this review. Research articles without peer review are not taken into consideration for review. There is no review of book chapters, patent applications, or communications for this review.

In this review, we are representing the statistics of the reviewed paper for different technologies. Fig. 1 represents the percentage of technology used in this literature survey: 48% digital twin's technology, 24 % & AR/VR, 14 % sensor & actuator, and 14 % IoT technology review for this review paper.

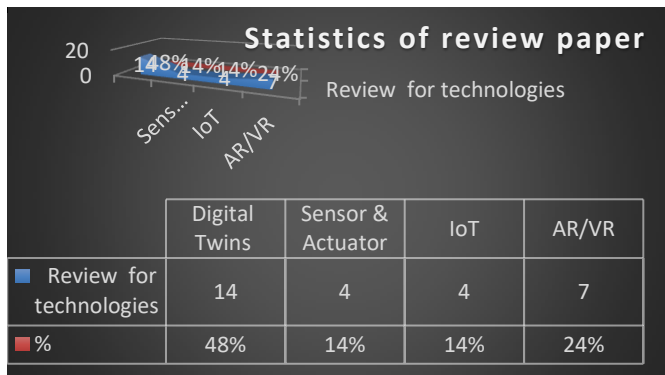


Fig. 1. Overview of the review paper of technologies.

## II. BACKGROUND OF THE HOSPITALITY INDUSTRY AND DIGITAL TWIN

Technologies used in the hospitality industry 4.0 present fresh possibilities for advancing sustainable growth [11]. Innovation has increased over the past few years in the hospitality business; self-registrations, contactless assistance, web-based seeking, and payment through applications have become the new norm [12]. While many businesses specifically target socioeconomics, the hospitality industry is more distinctive. Even inside the same place, Industry 4.0 offers diverse experiences to distinct groups [13]. Technologies used in the hospitality industry 4.0 present fresh possibilities for advancing sustainable growth. It can assist hospitality firms in streamlining operations; better marketing themselves, and meeting visitor needs [14] [15]. Organizations need to have fallback plans in place so they may be prepared for any challenges the world of innovation may throw at them [16]. If the innovation framework isn't set up at all or goes down, it can stop the entire presentation structure.

Customers give businesses online reviews via comments, ratings, and photos on internet platforms, which are growing in popularity daily. In order to improve its standing, the neighborhood company has been working hard to establish points of strength for interaction with customers [17]. Organizations can be destroyed or glorified by audits and comments; therefore, the business must take advantage of certain opportunities and manage its reputation. The hospitality industry is noted for having a high employee turnover rate, with about 33% of workers quitting after only six months on the job and about 45% remaining for an average of two years [18]. The industry is expected to continue to grow, which means that businesses must ensure that their representatives have strong personal qualities, professional skills, and knowledge in order to remain competitive [19]. Associations must be informed of the most recent trends for attracting and retaining employees because representative assumptions are constantly changing and evolving. This is becoming a constant challenge within the hospitality industry. Technology-based comparison is shown in Table I which shows the need for technology in the hospitality industry to improve the feature, quality, and services.

Online registration and looking voluntarily become more important to the business with visitors not coming to the front work area to receive a key due to new pleasant separation measures and a concentration on neatness [20]. As more people use their smartphones to request room management, computerized advancements made possible by mobile applications will become the norm for many businesses. The implementation of these innovations will require a significant initial investment but will ultimately result in efficiency and cheaper prices [21]. Rapid invention development has simultaneously created a variety of previously distinct arrangements that are now coming together.



TABLE I. OVERVIEW OF THE DIGITAL TWIN APPLICATION SCOPE AND FINDINGS

Ref.	Objective	Digital Twins (Sensor/Actuator/IoT)	Key Findings	Advantages	Future Scope
[7]	Hospitality Industry “Accommodation and services”	NA	Hospitality industry, Tourism, Services, Hospitality units, Tourists	Accommodation and service management are easy to improve for the customer	Technologies are required for customer satisfaction and service improvement
[8]	Identifying determinants of success in the development of new high-contact services: Insights from the hospitality industry	General Technologies such as mail and simple data collection methods are used only.	Service operations, Design and development, Hospitality services, Employee involvement	Services improvement, Market improvement, Process improvement and Hospitality improvement	Technologies are required to digitalize the hospitality industry for better customer satisfaction
[9]	Analyzing service quality in the hospitality industry	The general method used to make the hospitality industry reliable	Hospitality industry, Measurement, Service quality	Reliability, Responsiveness, Assurance, Empathy, Tangibles and Combined scale	Technologies are required to digitalize the hospitality industry for better customer satisfaction
[10]	Why do employees stay? a qualitative exploration of employee tenure	Quantitative and qualitative analysis of data without technology	Retention, Interviews, Turnover, Employee and Restaurant	Employee stability in hotels and restaurants, Hospitality Improvement	Technologies are required to digitalize the hospitality industry for a better customer Retention rate
[12]	Technology in Hospitality Industry	IoT, AR, Energy management, Beacon, and Automation	Interoperability, Data management, security, and Privacy	Modern Service platform for customer	Need to be overcome to institute a lasting, future proof solution for the hospitality industry.
[14]	The Digital Future of the Tourism & Hospitality Industry	AI, AR, VR and Blockchain	Customer Service, Customer Travelling, Technology used	Effectiveness, Improvement in Customer Services	Need to use the latest technologies to develop a sustainable and effective system for the hospitality industry
[40]	Research progress on virtual reality (VR) and augmented reality (AR) in tourism and hospitality	AI, AR	Augmented reality; Hospitality; Review; Technology; Tourism; Virtual reality	Virtually concept makes the easy to use for the customer in the hospitality industry	Need to use the latest technologies to develop a sustainable and effective system for the hospitality industry

This can create a challenging environment, especially for larger accommodation networks, which usually deal with a unique set of technological challenges [22]. However, integrating technologies and data across entire corporate ecosystems may be necessary to fully realize their potential [23]. Companies today use digital twin capabilities in several different ways. They are increasingly important mechanisms for modernizing entire manufacturing value chains and developing new goods in the hotel and aviation industries [24]. Operators in the hotel business collect and analyze massive amounts of in-hole data, which they then use to construct digital models that control drilling operations in real-time. With the use of digital twins, hotels, resorts, and other hospitality businesses can get an advantage over rival businesses, and then it is shown in Fig. 2. There are different elements, in which the digital twin can be beneficial in the hospitality industry such as interior direction; online concierge; personalized marketing; operation effectiveness.

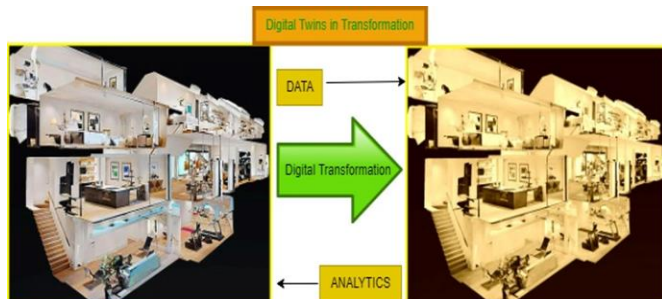


Fig. 2. Overview of the Digital Twins in the hospitality industry.

#### A. Interior Direction

The immediate benefit of employing digital twin technology in the hospitality sector is probably indoor navigation. Virtual representations of resorts, which are typically accessed through a mobile app, allow visitors to navigate to different on-site amenities like the pool or bar with turn-by-turn directions. Visitors can visually visit a hotel room or even the entire resort thanks to indoor navigation. Even when booking a suite, guests can specify a certain view or location. These amenities are a great way for prospective customers to choose whether they want to stay at a specific hotel.

#### B. Online Concierge

The level of customization and virtual concierge service that digital twins can provide are challenging to match. With a digital twin, visitors can remotely control anything from lighting and temperature to choosing a room from digital floor plans, having a drink delivered just to their location within the resort, and even submitting maintenance issues.

#### C. Personalized Marketing

Both guests and management benefit from this technology's targeted marketing advantages. Based on consumer behavior, management might promote certain room or service improvements to customers, encouraging guest spending. In turn, offerings that are pertinent, timely, and customized make guests happier.

#### D. Operation Effectiveness

The operational advantages that digital twins offer significantly enhance the guest experience. In order to save expenses (by optimizing staffing, lighting, and temperature during specific hours), comprehend how facilities are actually used, and meet operational performance goals, management might leverage data from consumer behavior. Finding possibilities to boost revenue, raise customer satisfaction, and maximize employee efficiency all depend on this data. Digital twins are advantageous to property managers as well as visitors. Through way finding, 3D visual experiences, virtual concierge services, and other means, they provide visitors with tailored experiences. They also give management analytical insight into how to improve staff productivity, increase operational efficiency, and deliver individualized guest marketing.

### III. OVERVIEW OF DIGITAL TWIN TECHNOLOGY

A digital twin is a representation of a physical product, procedure, or service in the digital world. A digital twin is a digital representation of a real-world object, such as a jet engine, wind farm, or even larger objects like a building or even an entire city [25]. The digital twin technology can be used to duplicate processes in order to gather data and forecast their performance, in addition to physical assets [26]. In essence, a digital twin is computer software that simulates how a process or product would work using data from the real world. To improve the output, these systems can use artificial intelligence, software analytics, and the internet of things [27]. Additionally, before any physical deployment is started, digital twin environments establish an environment that is conducive to testing new business operations, regulations, and assets to determine their performance levels [28].

The simple state to understand the digital twin is shown in Fig. 3. A virtual model can help identify surrenders and predict when an item's life will expire. Digital Twins can speed up creation, shorten the time it takes to market new products, and help reduce support expenses [29]. IoT, sensor, and actuator technologies are employed to convert physical systems into virtual concepts for novel setups and phenomena. This facilitates forecasting, adjusting, and decision-making on the real-time monitoring of performance and ease of future work [30].

Digital twin starts the work with sensor, actuator and IoT together to collect the information and to create the information to further digital work environment. There are two components of the digital twin to the work first is hardware component and other is software component [31]. Hardware components consist of IoT, sensor and actuator that assist the information for the whole process in the digital twins. In the software component the research engine, which turns naive observations into crucial business knowledge, is a key component of digital twinning. It is frequently governed by AI models [32]. In digital twin there are digital threat may be creating the complication in the data collection from the physical components of the system shown in Fig. 4. You can connect actual structures and their virtual representations into a closed circle known as a computerized string if all the necessary components are nearby.

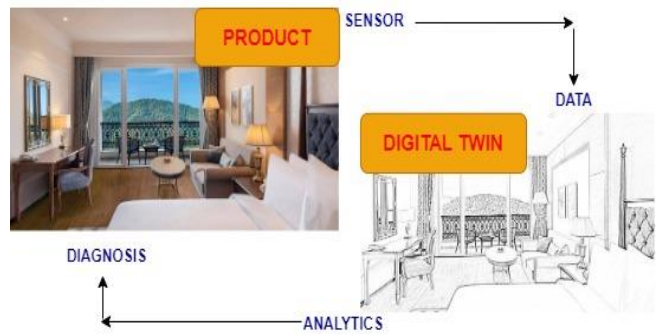


Fig. 3. Simple state and representation of digital twin concept.



Fig. 4. Digital threat in the software and hardware component of the digital twin.

Within the Digital threat, it involves some of the operations processes such as [33]: A physical object's environment at large is both sources of data that are transferred to the central repository. Data is produced for feeding to the digital threat after analysis. The digital twin uses new data to assess what would happen if the environment changes, discover bottlenecks, and mimic the object's operation in real-time. AI algorithms can be used at this stage to make product design adjustments, identify harmful tendencies, and avert expensive downtimes. The dashboard displays and visualizes analytics insights. Stakeholders make decisions based on actionable data. Accordingly, the physical object's parameters, procedures, or maintenance plans are modified [34].

Tandem Twinning Experts can evaluate the strength, adaptability, energy efficiency, and other characteristics of the various components that make up an item credit to the necessary degree of twinning [35]. To analyze how the part in question will behave under static or heated pressure and in other real-world circumstances, they can use reenactment programming.

Resource or item twinning the full item's replication reveals how various components work together under various conditions and how greater execution and dependability might be achieved. Instead of creating new models, advanced twinning can be used to develop new specialized arrangements [36]. This shortens the progression period and accounts for quicker emphasis.

Twining of the creation and cycle Advanced twinning applies to processes in addition to physical resources [37]. You create complete virtual models of the creation procedures for this scenario. This method helps to provide preferable answers to important questions like: How long will it take to produce a specific item? What will the price be? What should each machine accomplish? Which processes are automatable? Is there any way that a certain thing can be developed? Additionally, it is easier to avoid expensive free time when you can visualize the entire organizing process.

Device twinning Complex item and cycle interconnections and dependencies are made perceivable by a digital twin of the framework [38]. The twinned framework, which can be thought of as an arrangement of frames, can be practically as large as a multistory building, electrical lattice, or even an entire city. However, the risk involved in building such a reproduction typically does not equal the expected return. Because of this, framework twinning is typically not as flexible as other digital twin types.

#### IV. SENSOR AND ACTUATOR IN DIGITAL TWIN

Data from the physical system is collected by the digital twin and converted into digital data via the sensor and actuator [39]. The sensor and actuator are depicted in Fig. 5 and collaborate with IoT and digital twins. A sensor is a device that converts actual events or qualities into electrical signals [40]. This piece of technology converts the contribution that the weather makes and uses it to support the structure. As an illustration, a thermometer converts the temperature from a real sensor into electrical signals for the system. An actuator is a device that converts electrical signals into real-world events or characteristics.



Fig. 5. General representation of sensor and actuator.

Every sensor, including electromagnetic, capacitor, resistive, and others, has a different operating principle [41]. They often perceive the climate's contrasting quality and translate it into an electrical sign of corresponding size. A latent sensor doesn't require an additional power source to function, but a detached sensor does. A working piezoelectric sensor transforms strain into an electrical signal [42]. A potentiometer is an example of a detachable sensor since its resistance varies with location but needs additional power to convert it into an electrical signal.

#### V. IOT IN DIGITAL TWIN

Especially, the explosion in IoT sensors is crucial to the possibility of digital twins. Additionally, as IoT devices develop, advanced twin scenarios can involve less complex and modest products, providing additional benefits to businesses [43]. Digital twins can be used to predict different outcomes in light of varying knowledge. Computerized twins may usually enhance an IoT setup for maximum efficiency with additional programming and information analysis [44]. They can also help designers determine where things should go or how they should function before they are ever dispatched.

IoT is a technological revolution that represents the future of computing and communications, and its success is dependent on rapid technological advancement in a variety of sectors, from wireless sensors to nanotechnology [45]. It can essentially turn those items or appliances into 'smart' things that can transmit and receive data as well as communicate with one another. This can help with data collecting, automation, and allowing various devices to be managed or monitored from a single location, such as a phone or table. Its customization makes people feel special and since the primary aim of the hospitality industry revolves around providing the ultimate guest experience, this is what IoT should be embraced as shown in Fig. 6 [46]. While the concept of the IoT has been around for a long time, recent breakthroughs in a variety of technologies have made it a reality.

More liable technology: IoT innovation is turning out to be progressively open to additional producers on account of the accessibility of minimal expense, and high-dependability sensors.

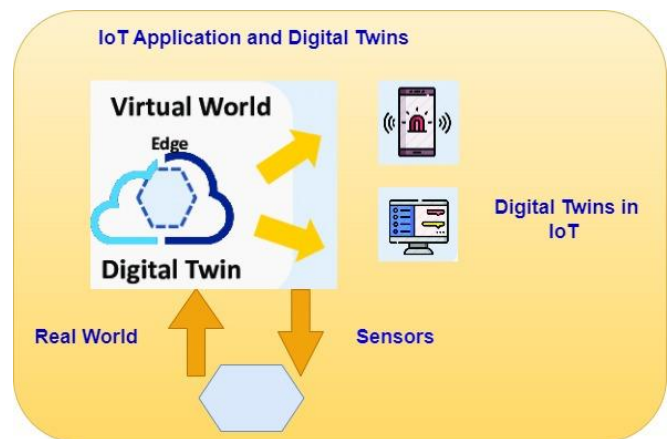


Fig. 6. General representation of IoT and its significance.

**Connectivity:** A variety of internet network protocols have made it simple to connect sensors to the cloud and other "things" for data transfer.

**Platform for Cloud computing:** As cloud platforms become more widely available, organizations and individuals can gain access to the infrastructure they need to scale up without having to manage it all.

**Analytics and Machine learning:** Organizations can obtain experiences quicker and all the more essentially on account of improvements in AI and examination, as well as admittance to different and huge volumes of information put away on the cloud. The development of these connected innovations keeps on pushing the outskirts of IoT, and IoT information takes care of these advances also.

**Conventional AI:** Natural language processing (NLP) has been brought to IoT gadgets (like computerized individual aides Alexa, Cortana, and Siri) because of advances in brain organizations, making them engaging, reasonable, and practical for home use.

## VI. AR AND VR IN DIGITAL TWIN

AR/VR has recently become a big idea in the hospitality industry since it allows hotels and other associated businesses to improve the actual environment they are offering or to enhance the experience of exploring the surrounding area [47]. AR is the perfect fusion of the real world and the electronic one to create a fake environment. Applications that use AR technology are developed for mobile devices or workspaces to integrate cutting-edge components into the current world [48]. A computer-produced reproduction of an alternate reality or environment is known as VR shown in Fig. 7. It is used in computer games and 3D movies [49].

AR shows the client relevant content by using computer vision, planning, and depth following. With the use of this functionality, cameras may collect, transmit, and interpret data to display cutting-edge material that is appropriate for the client being viewed [50] [51]. A VR headset screen must be placed in front of the client's eyes to remove any participation with our current reality in this way [52]. The built-in reality is vivid in augmented reality because you can also use visual, audible, and haptic excitement. AR and VR are used in hospitality management applications for various purposes [53].



Fig. 7. Virtual reality in hospitality industry.

## VII. DIGITAL TWIN IN HOSPITALITY INDUSTRY

A Digital twin is a representation of anything virtually [54]. This concept is evolving into an element of the dynamic process for increasing efficiency. In order to handle an item's nearly continuous state, operating condition, or position, digital twins use information from sensors that have been installed on the actual item [55]. This concept glorifies the replication processes of computer assisted design and computer assisted engineering. Any component of a physical object or process can be replicated using digital twins. The digital twin can reflect a new product's engineering drawings and measurements as well as all the subcomponents and associated lineage in the larger supply chain from the design table to the end user [56]. They might also appear in "as maintained" form, which would be a physical representation of the machinery on the factory floor.

The simulation depicts how the machinery works, engineers maintain it, or even how the consumer interacts with the products this machinery produces. Although digital twins can take many different forms, they all use and record data that simulates the real world. You can experiment with much iteration in the digital hotel sector to develop the scenarios that are most applicable to your company. New era industry 4.0 is the new revolution in hospitality industry to develop the simulation system on large scale to serve the society [57]. The hospitality industry has a wide range of uses for digital twins, which makes them a highly sought-after tool for investigating different market potential in this specialized area.

Real-time animation of the hospitality working process is also possible with the 3D plant model, which may be utilized to pinpoint the difficulties faced by this industry [58]. The primary factors that relate to the client for quality and service purposes are as illustrated in Fig. 8. Digital twin enables a virtual process to evaluate each step of the hospitality industry and provide an explanation of why a hotel and its services should be used [59]. The virtual facility gives a brief summary of all the features and amenities offered by the hotel as well as how the service provider addresses all the criteria in Fig. 7.. Over the coming years, it is anticipated that digital twin [60] applications will spread widely and no longer be restricted to activities or procedures exclusive [61] to the hotel business [62]. All types of hotel operations that want to stay competitive in their respective industries will use the technology. We can observe some instances of hospitality businesses that have already seized this chance. KFC Spain has joined MAPAL Data [63] Labs as the first business unit in the world to collaborate on a ground-breaking Digital Twin [64] project that uses cutting-edge digital [65] simulation to maximize labor and operational efficiencies [66]. It seems clear that digital twin technology will endure. The instrument has endless potential and offers numerous commercial advantages. Furthermore, although this technology [67] is currently viewed as a "nice-to-have," it will soon be necessary for companies that want to remain competitive and appealing to customers [68].

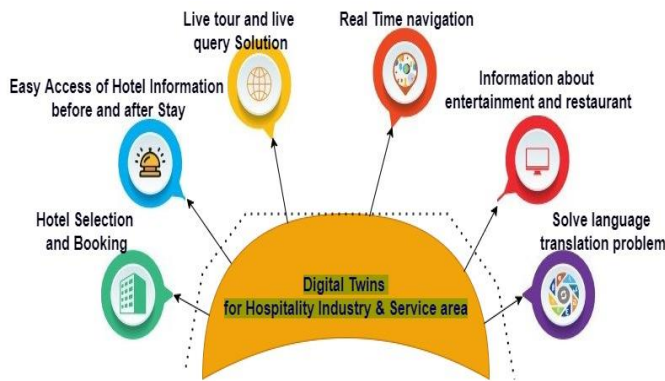


Fig. 8. Digital Twins for hospitality industry and simulation points.

### VIII. CONCLUSION

Digital twins play a crucial part in the technological advancement that has the potential to build a new foundation for the future. The merging of the physical and digital worlds makes it possible to make wise judgments at every stage of operations and hospitality, which can promote a data-driven smart hospitality environment. The use of the digital twin in the hospitality sector has drawn a lot of interest as a result of the facts mentioned above. In context of this purpose, the study's goal is to examine the value and potential applications of digital twins in the hospitality sector for the creation of cutting-edge digital infrastructure. The study also highlights many components that are important for the digital twin. The study concludes by summarizing and offering critical advice for the implementation of digital twins in the hospitality sector.

### IX. FUTURE WORK AND RECOMMENDATIONS

The research also examines technologies that benefit from and enable digital twinning. In-depth predictions are also included in the research for a variety of market areas and use cases, such as hotel service and simulations, production analytics, and others. A virtual object representation of a physical object that is mapped to actual objects in the real world, such as machinery, robots, or essentially any linked

business asset, is what is known as a "digital twin". IoT systems and software that are used to build a digital representation of the physical asset allow this mapping in the digital realm. A physical asset's digital twin can offer information about its status, including its physical state and disposition. The proposed digital architecture of the Hospitality Industry with digital twins is shown in Fig. 9.

On the other hand, tele-operation allows a digital object to be utilized to manipulate and control a real-world asset. This technology serves the hospitality industry in many ways such as:

- Future predictions from the digital twin solution: Planners can adjust for the following event iteration, improve operations, boost efficiency, and resolve any difficulties before they occur in a real-world setting by using digital twins, which frequently behave as a living, breathing model of the venue.
- Determine market obstacles and chances for digital twinning: By precisely recording their physical qualities, reproducing their actions, and altering their scale, a digital twin technology should be able to mimic both basic things and complex object relationships.
- Recognize the function of virtual twinning in product development, quality, and guest services: The ability to control quality and services is made possible by the vitality in digital twins, which will aid the hospitality business in the future. It makes the work faster to control and give the ability to handle the future challenges.
- Virtual simulations aid in understanding future plans and facilitate smooth decision-making with the least amount of money and effort: The hospitality industry can improve operational decision-making by utilizing the digital twin idea. Virtual reality has the functionality to help people make decisions more quickly in order to correct flaws.

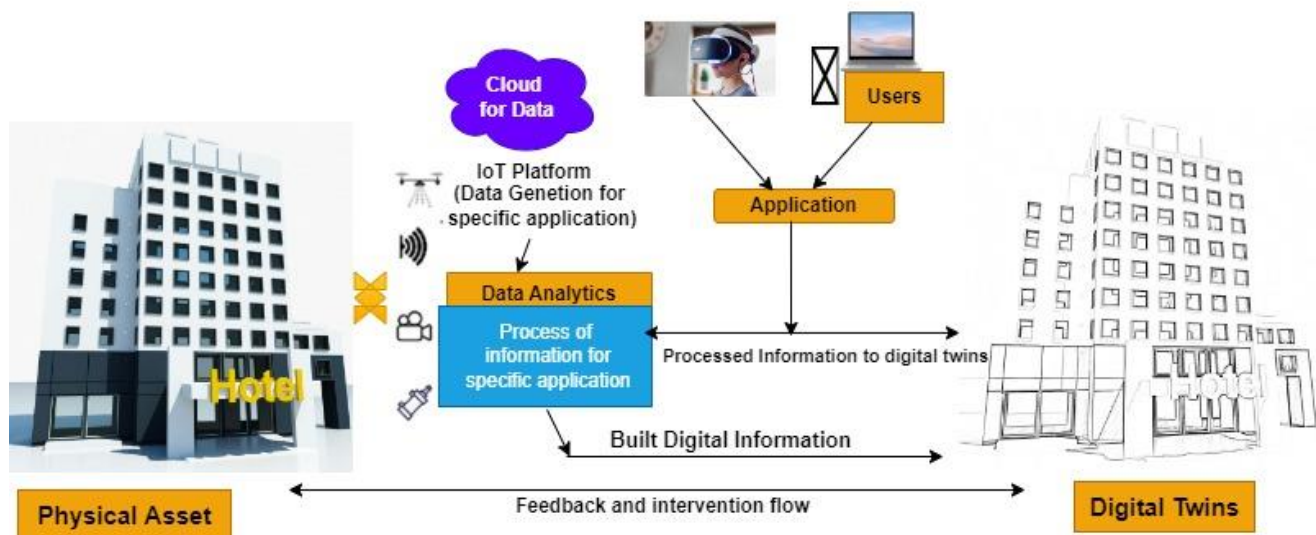


Fig. 9. Proposed architecture of the digital twin and hospitality industry.

REFERENCES

- [1] T. C. Liang and E. S. F. Wong, "Sustainable development: an adaptive re-use solution for the hospitality industry," *Worldwide Hospitality and Tourism Themes*, vol. 12, no. 5, pp. 623–637, 2020, doi: 10.1108/WHATT-06-2020-0047.
- [2] P. Jones, D. Hillier, and D. Comfort, "The Sustainable Development Goals and the Tourism and Hospitality Industry," *Athens Journal of Tourism*, vol. 4, no. 1, pp. 7–18, 2017, doi: 10.30958/ajt.4.1.1.
- [3] Aragón-correa, J. M. de Torre-ruiz, and M. D. Vidal-salazar, "Agglomerations around natural resources in the hospitality industry: Balancing growth with the sustainable development goals," 2022, doi: 10.1177/23409444221103283.
- [4] Neumann, K. Ott, and R. Kenchington, "Strong sustainability in coastal areas: a conceptual interpretation of SDG 14," *Sustainability Science*, vol. 12, no. 6, pp. 1019–1035, 2017, doi: 10.1007/s11625-017-0472-y.
- [5] X. Li, J. Cao, Z. Liu, and X. Luo, "Sustainable business model based on digital twin platform network: The inspiration from haier's case study in China," *Sustainability (Switzerland)*, vol. 12, no. 3, pp. 1–26, 2020, doi: 10.3390/su12030936.
- [6] K. T. Park et al., "Design and implementation of a digital twin application for a connected micro smart factory," *International Journal of Computer Integrated Manufacturing*, vol. 32, no. 6, pp. 596–614, 2019, doi: 10.1080/0951192X.2019.1599439.
- [7] J. Sun, Z. Tian, Y. Fu, J. Geng, and C. Liu, "Digital twins in human understanding: a deep learning-based method to recognize personality traits," *International Journal of Computer Integrated Manufacturing*, vol. 34, no. 7–8, pp. 860–873, 2021, doi: 10.1080/0951192X.2020.1757155.
- [8] W. Hu, T. Zhang, X. Deng, Z. Liu, and J. Tan, "Digital twin: a state-of-the-art review of its enabling technologies, applications and challenges," *Journal of Intelligent Manufacturing and Special Equipment*, vol. 2, no. 1, pp. 1–34, 2021, doi: 10.1108/jimse-12-2020-010.
- [9] V. Arrichiello and P. Gualeni, "Systems engineering and digital twin: a vision for the future of cruise ships design, production and operations," *International Journal on Interactive Design and Manufacturing*, vol. 14, no. 1, pp. 115–122, 2020, doi: 10.1007/s12008-019-00621-3.
- [10] S. M. Hasan, K. Lee, D. Moon, S. Kwon, S. Jinwoo, and S. Lee, "Augmented reality and digital twin system for interaction with construction machinery," *Journal of Asian Architecture and Building Engineering*, vol. 21, no. 2, pp. 564–574, 2022, doi: 10.1080/13467581.2020.1869557.
- [11] S. Shamim, S. Cang, H. Yu, and Y. Li, "Examining the feasibilities of Industry 4.0 for the hospitality sector with the lens of management practice," *Energies (Basel)*, vol. 10, no. 4, 2017, doi: 10.3390/en10040499.
- [12] M. Ionel, "Hospitality Industry," *Ovidius University Annals: Economic Sciences Series*, vol. 1, no. 1, pp. 187–191, 2016.
- [13] E. Bilotta, F. Bertacchini, L. Gabriele, S. Giglio, P. S. Pantano, and T. Romita, "Industry 4.0 technologies in tourism education: Nurturing students to think with technology," *Journal of Hospitality, Leisure, Sport and Tourism Education*, vol. 29, no. xxxx, p. 100275, 2021, doi: 10.1016/j.jhlste.2020.100275.
- [14] Issues and Sciences, "97C91De33529Cce4a4a07344E647E7B0E80," vol. 12, no. 3, 2019.
- [15] A. ben Youssef and A. Zeqiri, "Hospitality Industry 4.0 and Climate Change," *Circular Economy and Sustainability*, no. 0123456789, 2022, doi: 10.1007/s43615-021-00141-x.
- [16] M. Ottenbacher, J. Gnoth, and P. Jones, "Identifying determinants of success in development of new high-contact services: Insights from the hospitality industry," *International Journal of Service Industry Management*, vol. 17, no. 4, pp. 344–363, 2006, doi: 10.1108/09564230610680659.
- [17] M. O. W. Amy, D. A. M., and C. J. White, "AnalisisPenerapanStandarPelayanan Minimal di BidangKesehatanpadaIndikatorPelayananKesehatanPenderitaHipertensi diPuskesmas Kota Semarang," *FakultasKesehatanMasyarakat, UniversitasDiponegoro*, 1999.
- [18] T. Self and B. Dewald, "Why do employees stay? a qualitative exploration of employee tenure," *International Journal of Hospitality and Tourism Administration*, vol. 12, no. 1, pp. 60–72, 2011, doi: 10.1080/15256480.2011.540982.
- [19] S. K. Hight, T. Gajjar, and F. Okumus, "Managers from 'Hell' in the hospitality industry: How do hospitality employees profile bad managers?," *International Journal of Hospitality Management*, vol. 77, no. February, pp. 97–107, 2019, doi: 10.1016/j.ijhm.2018.06.018.
- [20] P. Kansakar, A. Munir, and N. Shabani, "Technology in the Hospitality Industry: Prospects and Challenges," *IEEE Consumer Electronics Magazine*, vol. 8, no. 3, pp. 60–65, 2019, doi: 10.1109/MCE.2019.2892245.
- [21] M.-A. Popescu, F.-V. Nicolae, and M.-I. Pavel, "Tourism and Hospitality Industry in the Digital Era: General Overview," *Proceedings of the 9Th International Management Conference: Management and Innovation for Competitive Advantage*, pp. 163–168, 2015, [Online]. Available: www.internetlivestats.com,
- [22] Martin Zsarnoczky, "The Digital Future of the Tourism & Hospitality Industry By Martin Zsarnoczky Spring 2018," *Boston Hospitality Review*, vol. Spring, no. June, pp. 1–9, 2018, [Online]. Available: www.bu.edu/bhr
- [23] F. Psarommatis and G. May, "A literature review and design methodology for digital twins in the era of zero defect manufacturing," 2022, doi: 10.1080/00207543.2022.2101960.
- [24] Zhang and G. Y. Tian, "UHF RFID Tag Antenna-Based Sensing for Corrosion Detection & Characterization Using Principal Component Analysis," *IEEE Transactions on Antennas and Propagation*, vol. 64, no. 10, pp. 4405–4414, 2016, doi: 10.1109/TAP.2016.2596898.
- [25] A. el Saddik, "Digital Twins: The Convergence of Multimedia Technologies," *IEEE Multimedia*, vol. 25, no. 2, pp. 87–92, 2018, doi: 10.1109/MMUL.2018.023121167.
- [26] R. Rosen, G. von Wichert, G. Lo, and K. D. Bettenhausen, "About the importance of autonomy and digital twins for the future of manufacturing," *IFAC-PapersOnLine*, vol. 28, no. 3, pp. 567–572, 2015, doi: 10.1016/j.ifacol.2015.06.141.
- [27] A. S. Duggal et al., "A sequential roadmap to Industry 6.0: Exploring future manufacturing trends," *IET Communications*, vol. 16, no. 5, pp. 521–531, 2022, doi: 10.1049/cmu2.12284.
- [28] E. Harper, C. Ganz, and K. E. Harper, "Digital Twin Architecture and Standards," *IIC Journal of Innovation*, no. November, pp. 1–12, 2019.
- [29] K. M. Alam and A. el Saddik, "C2PS: A digital twin architecture reference model for the cloud-based cyber-physical systems," *IEEE Access*, vol. 5, pp. 2050–2062, 2017, doi: 10.1109/ACCESS.2017.2657006.
- [30] A. K. Ghosh, A. S. Ullah, R. Teti, and A. Kubo, "Developing sensor signal-based digital twins for intelligent machine tools," *J IndIntegr*, vol. 24, p. 100242, 2021, doi: 10.1016/j.jii.2021.100242.
- [31] G. Kapteyn, D. J. Knezevic, D. B. P. Huynh, M. Tran, and K. E. Willcox, "Data-driven physics-based digital twins via a library of component-based reduced-order models," *International Journal for Numerical Methods in Engineering*, pp. 1–18, 2020, doi: 10.1002/nme.6423.
- [32] G. Kapteyn, D. J. Knezevic, and K. E. Willcox, "Toward predictive digital twins via component-based reduced-order models and interpretable machine learning," *AIAA Scitech 2020 Forum*, vol. 1 PartF, no. January, pp. 1–19, 2020, doi: 10.2514/6.2020-0418.
- [33] Mazak, S. Wolny, and M. Wimmer, *On the Need for Data-Based Model-Driven Engineering*. 2019. doi: 10.1007/978-3-030-25312-7\_5.
- [34] K. Alshammari, T. Beach, and Y. Rezgui, "Cybersecurity for digital twins in the built environment: Current research and future directions," *Journal of Information Technology in Construction*, vol. 26, no. May 2020, pp. 159–173, 2021, doi: 10.36680/j.itcon.2021.010.
- [35] Y. Lu, C. Liu, K. I. K. Wang, H. Huang, and X. Xu, "Digital Twin-driven smart manufacturing: Connotation, reference model, applications and research issues," *Robotics and Computer-Integrated Manufacturing*, vol. 61, no. July 2019, p. 101837, 2020, doi: 10.1016/j.rcim.2019.101837.
- [36] J. Leng et al., "Digital twin-driven rapid reconfiguration of the automated manufacturing system via an open architecture model,"

- Robotics and Computer-Integrated Manufacturing, vol. 63, no. March 2019, 2020, doi: 10.1016/j.rcim.2019.101895.
- [37] S. Reed, M. Löfstrand, and J. Andrews, "Modelling cycle for simulation digital twins," *Manufacturing Letters*, vol. 28, pp. 54–58, 2021, doi: 10.1016/j.mfglet.2021.04.004.
- [38] J. Leng, H. Zhang, D. Yan, Q. Liu, X. Chen, and D. Zhang, "Digital twin-driven manufacturing cyber-physical system for parallel controlling of smart workshop," *Journal of Ambient Intelligence and Humanized Computing*, vol. 10, no. 3, pp. 1155–1166, 2019, doi: 10.1007/s12652-018-0881-5.
- [39] S. T. Smith and R. M. Seugling, "Sensor and actuator considerations for precision, small machines," *Precision Engineering*, vol. 30, no. 3, pp. 245–264, 2006, doi: 10.1016/j.precisioneng.2005.10.003.
- [40] Hać and L. Liu, "Sensor and actuator location in motion control of flexible structures," *Journal of Sound and Vibration*, vol. 167, no. 2, pp. 239–261, 1993, doi: 10.1006/jsvi.1993.1333.
- [41] J. A. Stankovic, "When sensor and actuator networks cover the world," *ETRI Journal*, vol. 30, no. 5, pp. 627–633, 2008, doi: 10.4218/etrij.08.1308.0099.
- [42] E. Y. Song, M. Burns, A. Pandey, and T. Roth, "IEEE 1451 Smart Sensor Digital Twin Federation for IoT/CPS Research," *SAS 2019 - 2019 IEEE Sensors Applications Symposium, Conference Proceedings*, pp. 1–6, 2019, doi: 10.1109/SAS.2019.8706111.
- [43] A. Simchenko, S. Y. Tsohla, and P. P. Chyvatkin, "IoT & digital twins concept integration effects on supply chain strategy: Challenges and effect," *International Journal of Supply Chain Management*, vol. 8, no. 6, pp. 803–808, 2019.
- [44] V. Kamath, J. Morgan, and M. I. Ali, "Industrial IoT and Digital Twins for a Smart Factory: An open source toolkit for application design and benchmarking," *GIOTS 2020 - Global Internet of Things Summit, Proceedings*, pp. 0–5, 2020, doi: 10.1109/GIOTS49054.2020.9119497.
- [45] S. G. H. Soumyalatha, "Study of IoT: Understanding IoT Architecture, Applications, Issues and Challenges," *International Journal of Advanced Networking & Applications (IJANA)*, no. May 2016, pp. 1–5, 2019.
- [46] S. H. Shah and I. Yaqoob, "A survey: Internet of Things (IoT) technologies, applications and challenges," *2016 4th IEEE International Conference on Smart Energy Grid Engineering, SEGE 2016*, vol. i, pp. 381–385, 2016, doi: 10.1109/SEGE.2016.7589556.
- [47] R. Yung and C. Khoo-Lattimore, "New realities: a systematic literature review on virtual reality and augmented reality in tourism research," *Current Issues in Tourism*, vol. 22, no. 17, pp. 2056–2081, 2019, doi: 10.1080/13683500.2017.1417359.
- [48] W. Wei, "Research progress on virtual reality (VR) and augmented reality (AR) in tourism and hospitality: A critical review of publications from 2000 to 2018," *Journal of Hospitality and Tourism Technology*, vol. 10, no. 4, pp. 539–570, 2019, doi: 10.1108/JHTT-04-2018-0030.
- [49] Bec, B. Moyle, V. Schaffer, and K. Timms, "Virtual reality and mixed reality for second chance tourism," *Tourism Management*, vol. 83, no. November 2020, p. 104256, 2021, doi: 10.1016/j.tourman.2020.104256.
- [50] M. Billinghurst, A. Clark, and G. Lee, "A survey of augmented reality," *Foundations and Trends in Human-Computer Interaction*, vol. 8, no. 2–3, pp. 73–272, 2014, doi: 10.1561/1100000049.
- [51] J. Carmigniani and B. Furht, *Handbook of Augmented Reality*. 2011, doi: 10.1007/978-1-4614-0064-6.
- [52] F. Biocca, "Virtual Reality Technology: A Tutorial," *Journal of Communication*, vol. 42, no. 4, pp. 23–72, 1992, doi: 10.1111/j.1460-2466.1992.tb00811.x.
- [53] J. M. Zheng, K. W. Chan, and I. Gibson, "Virtual reality," *IEEE Potentials*, vol. 17, no. 2, pp. 20–23, 1998, doi: 10.1109/45.666641.
- [54] N. S. Dang, H. Kang, S. Lon, and C. S. Shim, "3D digital twin models for bridge maintenance," *Proceedings of 10th International Conference on Short and Medium Span Bridges*, no. 73, pp. 1–9, 2018, [Online]. Available: [https://www.researchgate.net/publication/331314334%0Ahttps://www.csee.ca/elf/apps/CONFERENCEVIEWER/conferences/SMSB/papers/FinalPaper\\_73\\_0508011616.doc](https://www.researchgate.net/publication/331314334%0Ahttps://www.csee.ca/elf/apps/CONFERENCEVIEWER/conferences/SMSB/papers/FinalPaper_73_0508011616.doc)
- [55] J. An, C. Kai Chua, and V. Mironov, "Application of Machine Learning in 3D Bioprinting: Focus on Development of Big Data and Digital Twin," 2021, doi: 10.18063/ijb.v7i1.342.
- [56] M. Singh, E. Fuenmayor, E. P. Hinchy, Y. Qiao, N. Murray, and D. Devine, "Digital twin: Origin to future," *Applied System Innovation*, vol. 4, no. 2, pp. 1–19, 2021, doi: 10.3390/asi4020036.
- [57] F. Pires, A. Cachada, J. Barbosa, A. P. Moreira, and P. Leitao, "Digital twin in industry 4.0: Technologies, applications and challenges," *IEEE International Conference on Industrial Informatics (INDIN)*, vol. 2019-July, pp. 721–726, 2019, doi: 10.1109/INDIN41052.2019.8972134.
- [58] Z. Zhu, C. Liu, and X. Xu, "Visualisation of the digital twin data in manufacturing by using augmented reality," *Procedia CIRP*, vol. 81, pp. 898–903, 2019, doi: 10.1016/j.procir.2019.03.223.
- [59] Raj and C. Surianarayanan, *Digital twin: The industry use cases*, 1st ed., vol. 117, no. 1. Elsevier Inc., 2020, doi: 10.1016/bs.adcom.2019.09.006.
- [60] Kampker, V. Stich, P. Jussen, B. Moser, and J. Kuntz, "Business models for industrial smart services - the example of a digital twin for a product-service-system for potato harvesting," *Procedia CIRP*, vol. 83, pp. 534–540, 2019, doi: 10.1016/j.procir.2019.04.114.
- [61] "KFC Spain reduced labour cost by 2.65% with the help of MAPAL Workforce - MAPAL OS." <https://mapal-os.com/en/resources/success-stories/kfc-spain> (accessed Aug. 09, 2022).
- [62] Bharany, S., Sharma, S., Frnda, J., Shuaib, M., Khalid, M. I., Hussain, S., Iqbal, J., & Ullah, S. S. (2022). Wildfire Monitoring Based on Energy Efficient Clustering Approach for FANETS. In *Drones* (Vol. 6, Issue 8, p. 193). MDPI AG. <https://doi.org/10.3390/drones6080193>
- [63] Gehlot, A., Singh, R., Kathuria, S., Chhabra, G., & Joshi, K. (2023, March). Cloud based E-Feedback System for Hospitality Industry. In *2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)* (pp. 1438-1442). IEEE.
- [64] Bharany, S., Sharma, S., Alsharabi, N., Tag Eldin, E., & Ghamry, N. A. (2023). Energy-efficient clustering protocol for underwater wireless sensor networks using optimized glowworm swarm optimization. In *Frontiers in Marine Science* (Vol. 10). Frontiers Media SA. <https://doi.org/10.3389/fmars.2023.1117787>
- [65] Raman, R., Joshi, K., Kumar, G. S., Ramachandran, K. K., Bothe, S., & Trivedi, S. (2023, May). Benefits of Implementing an Ad-Hoc Network for Hospitality Businesses with IoT Smart Devices. In *2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)* (pp. 2042-2046). IEEE.
- [66] Bharany, S., Badotra, S., Sharma, S., Rani, S., Alazab, M., Jhaveri, R. H., & Reddy Gadekallu, T. (2022). Energy efficient fault tolerance techniques in green cloud computing: A systematic survey and taxonomy. In *Sustainable Energy Technologies and Assessments* (Vol. 53, p. 102613). Elsevier BV. <https://doi.org/10.1016/j.seta.2022.102613>
- [67] Tayal, P., Rastogi, N., Ahuja, T. K., Tyagi, S., Joshi, K., & Mohialden, Y. M. (2022, November). Impact Of Ai On The Banking Industry 4.0. In *2022 7th International Conference on Computing, Communication and Security (ICCCS)* (pp. 1-4). IEEE.
- [68] Joshi, K., Anandaram, H., Khanduja, M., Kumar, R., Saini, V., & Mohialden, Y. M. (2022). Recent Challenges on Edge AI with Its Application: A Brief Introduction. In *Explainable Edge AI: A Futuristic Computing Perspective* (pp. 73-88). Cham: Springer International Publishing.

# Design of a Hypermodel using Transfer Learning to Detect DDoS Attacks in the Cloud Security

Marram Amitha, Dr.Muktevi Srivenkatesh

Department of Computer Science, GITAM Deemed to be University, Visakhapatnam, India

**Abstract**—The present research proposes a detective approach to analyzing the performance of various algorithms used for more accurate detection of Distributed Denial-of-Service (DDoS) attacks in cloud computing. From the start, this study uses machine learning and deep learning to explore whether information security has evolved in recent years. The deployment of intrusion detection systems and distributed denial-of-service attacks are then discussed. The most common DDoS attack types were summarized. In addition, this study reviewed the existing approaches and techniques for DDoS attack detection. Various pre-processing subsystems as well as attribute-based selection techniques for preventing the detection of DDoS were briefly described. The proposed Intrusion detection system uses transfer learning for detecting DDoS attacks in the Networks. The proposed system used for the data set for the Network Intrusion Detection System is SDN Dataset which has more features and is suitable to use to detect in Network Intrusions. It contains 23 features that are used to detect intrusions in the network SDN Dataset which consists of training and testing data to detect the attacks in the network. The detection and prevention subsystems through ML and DL strategies were briefly discussed. The proposed deep learning model for DDoS attack detection in cloud storage applications is explained. After that, various preprocessing strategies employed in the detection are described, among them rebalancing data, data cleaning, data splitting, and data normalization like min-max normalization. The author created a hypermodel that consists the parameters of baseline classifiers like Support Vector Machine, K-Nearest Neighbors Algorithm, XGboost, and other various machine learning models. The proposed model gives very good accuracy compared to other machine learning models.

**Keywords**—Machine learning; deep learning; support vector machine; k-nearest neighbors algorithm

## I. INTRODUCTION

Intrusion relates to every collection of connected operations performed by a malevolent adversary that affects a target system. Regarding the detection of DoS and DDoS attacks, intrusive activities are usually adaptable and can be categorized based on the attacks.

The primary threat from these four intrusion activities involves a DoS and DDoS attack that either consumes computer and communication facilities or takes advantage of the system's vulnerabilities to make the system accessible for authorized users. This leads to a significant loss of resources, money, and data. When numerous systems overwhelm the internet connection of a targeted system, which includes one or more web servers [1], a Distributed Denial-of-Service (DDoS) assault occur. Such an attack usually arises with traffic

overloading of the targeted system resulting from many compromised systems. Distributed Denial of Service attacks are different from different kinds of attacks in that can carry conduct a damaging attack on Internet-connected resources.

### A. Phases of DDoS Attack

Two phases are followed in DDoS attacks. The attacker seeks to gain access to the network's vulnerable machines first. With the compromised hosts of other networks, the attacker or master establishes his own network and describes them as slaves. The 'intrusion phase' occurs when that occurs. The attacker then selects which victim server to target and starts delivering packets in that direction. With DoS attacks, which start from a single host, DDoS attacks originate from a number of dynamic networks that were previously hacked. The 'DDoS attack phase' [2] corresponds to what this is recognized as.

Based on the way methods work, DoS and DDoS attacks can be roughly classified into three categories (Ali et al., 2019). These involve assaults based on connection utilization, bandwidth consumption, and vulnerability exploitation.

### B. Connection Consumption-based Attacks

A connection-oriented protocol is the Transmission Control Protocol (TCP). In advance of data exchange, it establishes a connection between the client and the server. A limited quantity of connection requests can be accepted and processed by any server. In an effort keep those with authorization considering accessing the service offered by the organization, the attacker establishes an enormous amount of connections with the server. The operating system's kernel [5] resources required for setting up connections have been drained by this type of attack. One of the most frequently used attacks under this category is the SYN Flood attack.

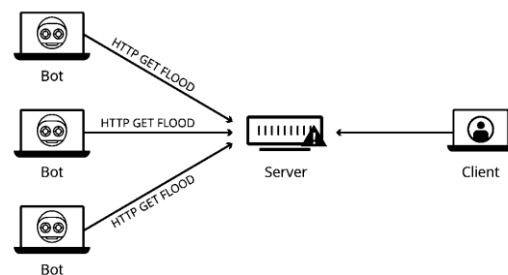


Fig. 1. Flood attack.

Fig. 1 shows the process of a SYN Flood attack, during which the attacker establishes an extensive amount of TCP connections with connections only partially accessible to deplete the connection pool. Due to how it operates, a cluster



of servers may be configured to function slower than normal through using slow network connections.

### C. Bandwidth Consumption-based Attacks

Every network has an established amount of accessible bandwidth. The network's bandwidth limit has been exceeded, which will impact up the response times of the servers and the various devices linked to it. Attacks based on computation-based DoS and DDoS are started by making use of this primary bandwidth. In order to generate a massive flood, the attacker uses established handler machines to control an enormous quantity of preconfigured zombie devices connected to the internet. User Datagram Protocol (UDP) flood is a prevalent attack in this type of attack. Fig. 2 shows the visualization of continuous features with respect to packet count protocol and type of attack.

This paper explains the procedure of transfer learning to train a model to get higher accuracy. In this research, authors used A.DDoS attack SDN Datasets from Mendeley website.

## II. RELATED WORK

Deep learning is gaining popularity these days because of its higher accuracy and performance. Its implementations in this field are being researched by a community of scholars. Automotive architecture, healthcare, manufacturing, and law enforcement are some of the well-known realms. The study that has already been completed by various scholars is mentioned below. Asad et al. [2020] [1] equate the effects of the machine learning methodology to those of others The Naive Bayes categorization methodology and the decision-tree categorization methodology are two examples of machine learning techniques. The researcher mostly attempted to act in a version that was not online. If the scale of the dataset grows larger, the output disparity becomes more pronounced. Deep Intelligence was introduced by Bhuvanewari Amma N.G et al [2023] [2]. The knowledge was derived using a radial base function with a variety of abstraction levels. The research was conducted on well-known datasets such as NSL-KDD and UNSW NB15 that included 27 functions. In comparison to other existing methods, the researcher believed that his method was more accurate. Muhammad Aamir et al. used a clustering technique to apply a feature selection process. Five related machine-learning methodologies were used to compare the algorithm. For preparation, SVM and RF methodologies were utilized. The best accuracy was attained by RF, which was about 96 percent.

Mishra et al. [3] classified packages depending on their characteristics. Through inspecting the IP header, the protection strategy attempts to identify IP addresses. These IP addresses are utilized to distinguish between spoofed and legitimate addresses. As the scale of the assault becomes larger, firewalls are ineffective. For separating the regular and assaulted traffic, Narasimha et al utilize anomaly identification and machine learning methodologies. Real-time datasets were utilized in the research. For classification, the well-known naive Bayes ML methodology was utilized. The outcomes were compared to those of other methodologies such as J48 and RF.

A. Haddaji et al [2023] [4] combined intellectual-stimulated computation and the entropy methodology in their research. For the classification, Support Vector Machine Learning was utilized. The platform's flow chart was being mined for information. In terms of identification precision, the outcomes were satisfactory. Omar E. Elejla et al. used an IPv6 classification strategy to incorporate a methodology for the identification of DDoS assaults. The findings were compared to 5 different well-known machine learning methodologies by the scientist. DT, SVM, NB, KNN, and NN were the methodologies utilized. The research was conducted on a well-known dataset. According to the source, the KNN methodology achieved a precision of about 85 percent.

Farhan ulla et al. [2023] [5] used the ML methodology to create an entropy-depend semi-supervised methodology. Unsupervised and supervised designs are used in this development. Unsupervised techniques have high precision and low false-positive rates. Supervised methods, contrarily, minimize the number of false positives. The datasets utilized in the research were NSL-KDD, UNB ISCX 12, and UNSW-NB15. For the recognition of the assault, Nathan Shone et al. use a DL methodology. It also utilized the NDAE function for unsupervised instruction. On the well-known KDD Cup 99 and NSL-KDD datasets, the proposed methodology was executed on a GPU utilizing TensorFlow. The researcher believed that he was able to get more precise identification outcomes.

## III. INTRUSION DETECTION SYSTEM MODEL

The Internet is a global network of computers connected through various media and a standard protocol. Among many additional essential elements in modern life, people nowadays depend on the World Wide Web for their education, trade, social ability, and recreational activities. Evidently, the Internet brought perhaps the greatest advances in communication and computing.

Attacks on the web may involve in many different possible dangers, such as financial loss, identity theft, loss of confidential data or information, theft of network resources, damage to a person's brand and reputation, and a decrease in consumer confidence in online banking and e-commerce.

Most security issues differentiate themselves from their earlier equivalents in non-network infrastructures since data and business logic are located on a remote Network server with transparent supervisors. One of these attacks, the Denial of Service (DDoS) attack, has been extremely aggressive and extremely intrusive to web servers. Denial of Service (DDoS) assaults frequently target the server a network of computers that provide consumers a service. DoS attackers seek to consume servers that are operational in an approach that leads the service stop functioning because of an excessive number of outstanding requests in the service queue.

DDoS attacks can be performed on government departments, educational institutions, and home systems that have been hacked. These computer programs are referred to as bots. Usually, DoS assaults begin at the network layer by sending many UDP, SYN, or ICMP packets of data. Attackers migrate to the application layer and flood it with HTTP GET requests, which is referred to as application-layer DDoS, after

network layer attack fails. According to Bhardwaj et al. (2020) [6], DDoS attacks can use TCP SYN, UDP flood, DNS reflection, HTTP flood, and ICMP flood. Based on information research on security, distributed denial-of-service (DDoS) attacks recently cost companies and governments all through the world a large amount of revenue. On the other side, the attackers use more advanced techniques to amplify attacks and overload their targets by taking benefit of their geographic distribution and computing power provided possibly by the wide variety of devices and their various movements, which are frequently incorporated within IoT network scenarios. Therefore, a practical and effective DDoS detection method must be developed and has been optimized, and functional in IoT-based smart environments with major constraints on processor power, reaction time, and data processing volume. Thus, conventional IDSs might out to be completely appropriate for applications in the Internet of Things. IoT security is a continuously significant issue that requires the creation of mitigation techniques and an increasing understanding of IoT safety concerns system (Catak FO et al., 2020) [7]. In order to guarantee that client data is stored in a safe fashion, security is another crucial component of network-based IoT data. Network security is crucial for both commercial and personal users. Everyone wants to remain certain regarding the integrity of their personal data. Businesses are legally required to protect customer information, and certain industries require more stringent regulations around data storage. Various difficulties with the value multi-tenancy, data loss and leaking, network accessibility, identity management, harmful APIs, inconsistent service levels, patch management, and internal threats are associated with the issue of network computing security. Some safety features that cannot be sufficiently scalable, incompatible, and appropriate are missing from conventional basic cryptographic algorithms. With these requirements in mind, an IDS mechanism with the indicated encryption approach has been established to defend the network from DDoS attacks.

#### A. DDoS Attack SDN Dataset

Machine learning and deep learning algorithms use this smaller net emulator-generated data set that has been modified with SDN, to categorize traffic. At the start of the project, ten smaller networks with switches connected to a single Ryu controller are established. Network simulation was utilized for imitating malicious traffic, such as TCP Syn assaults, UDP flood incidents, and ICMP assaults, as well as normal traffic including TCP, UDP, and ICMP. A total of 23 features in everything, some of which have been determined and some of which were obtained from the switches, in the data set. Among of the characteristics that have been obtained are Switch-id, Packet\_count, Byte\_Count, Duration\_sec, Duration\_nsec, which is Duration in Nanoseconds, Source IP, Destination IP, and Total Duration. The port symbol while rx\_bytes indicates the number of bytes received on the switch port, tx\_bytes indicates the number of bytes carried from the switch port. The dt field shows the time and date prior to the are transferred to values, and a flow is tracked every 30 seconds. Packet per flow examines the entire amount of packet in a single flow, while bytes per flow examines the total amount of bytes in a single flow. The data transfer and reception accelerates are tx\_kbps

and rx\_kbps, respectively, whereas port bandwidth is the product of tx\_kbps and rx\_kbps. The packet rate, which is determined as the number of packets sent per second, can be determined by dividing the number of packets provided per flow by the monitoring interval, Packetins message measure, and the total number of flow entries in the switch. The last column's category proof of identity, which decides whether the traffic type is malicious, is shown. Label 1 indicates malicious traffic, while Label 0 indicates benign traffic. The result of a 250-minute network simulation produced 1,04,345 rows of statistics. Repeating the simulation for a longer amount of time allows for the gathering of more data.

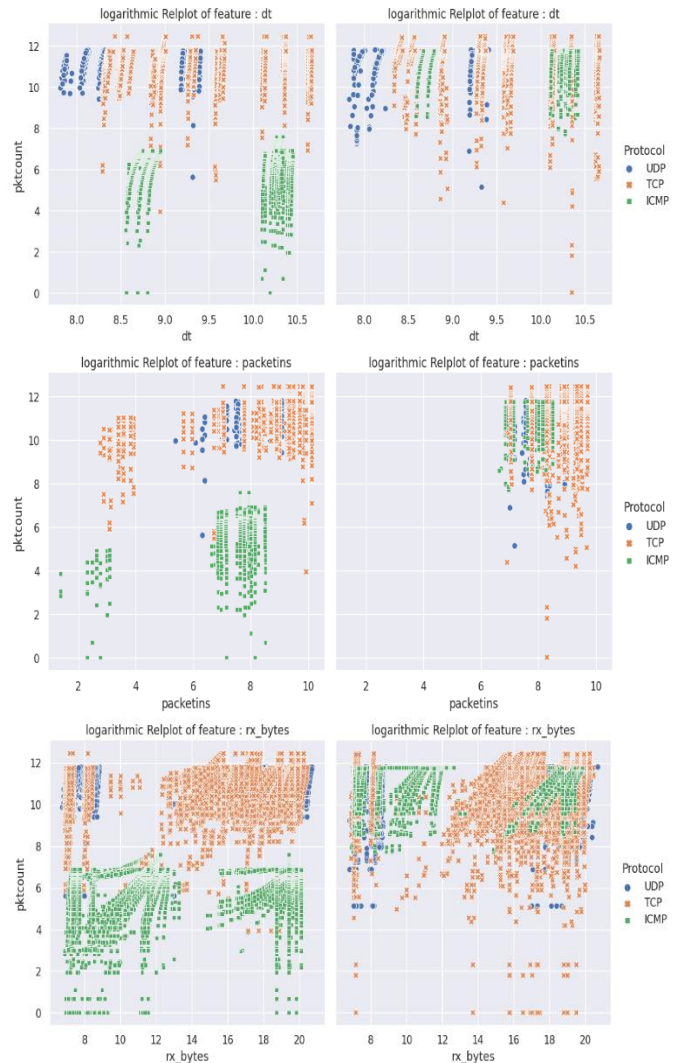


Fig. 2. Visualization of continuous features with respect to packet count, protocol, and type of attack.

#### IV. PROPOSED METHODOLOGY

A sophisticated model based on DLNN which employs optimized values in the hidden layers to differentiate among normal and attacked network data is developed for recognizing DDoS attacks.

### A. Deep Learning Neural Network

One of the most advanced artificial intelligence methods for tackling computer vision's 11 challenges is the Deep Convolutional Neural Network (Deep CNN). As a feed-forward artificial neural network, the Deep CNN established a class of deep learning and has been utilized for different agricultural image classification research. Deep CNN's convolutional layer, which is vital, employs filters to extract data from the input images. An enormous amount of training data must be collected for the purpose of enhancing the performance of Deep CNN. Fig. 3 shows the architecture of the Deep CNN technique.

One of the primary advantages of using Deep CNN for image classification is avoiding the requirement for feature development. Deep CNN's numerous levels each contain multiple convolutions. They provide many kinds of visuals for the training data in the quicker, more detailed layers, acquiring to more intricate ones in the deeper layers. The pooling layers, which initially serve as methods to extract features from the convolutional layers' performance as feature extractors, are then used to decrease the dimensionality of the training data. In the words of Chen J et al. (2019) [8], the convolutional layers transform an assortment of lower-level features into additional discriminative features. The vital elements of Deep CNN are the convolutional layers in addition. In contrast to traditional machine learning, feature engineering is an essential part of deep learning. The down-sampling process gets carried out along the spatial dimensions through the pooling layer. It promotes having fewer parameter choices. The pooling component of the proposed model uses the max-pooling process. In the proposed Deep CNN model, max pooling outperforms average pooling in processing performance. Dropout, which explains removing entities from the network, is another important layer. It follows the overfitting reducing regularization strategy. Using dropout values that ranged from 0.2 to 0.8, the proposed model was trained and compared. Applying the convolutional and pooling layers results, the dense layer follows up with the classification.

Deep CNN is a highly iterative process that requires developing an assortment of models while deciding on the most effective one (Morgan Kaufmann, et al. 2019) [9].

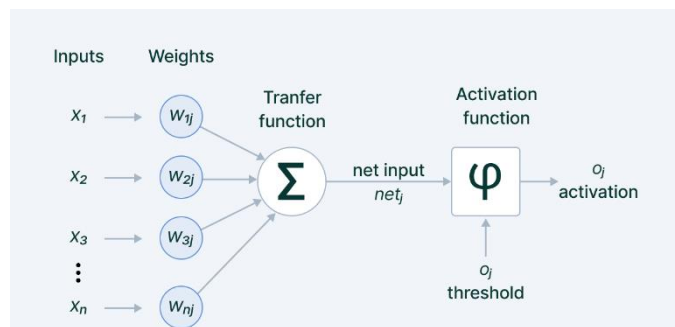


Fig. 3. The architecture of the deep CNN technique.

### B. Transfer Learning

An approach for transmitting knowledge from one machine learning model to another is referred to as transfer learning by collecting bias and weight values from present models, it

reduces the initial model construction phase of the new model. For instance, a machine learning model developed for task A could be used as a foundation for a model for task B. Current pre-trained models are used through transfer learning to gather knowledge that can be applied to new models.

In the research of Panigrahi R et al. (2018), AlexNet, Visual Geometry Group Network (VGGNet), Residual Network (ResNet), and Inception Network are among the most frequently utilized pre-trained deep learning models. A popular pre-trained deep convolutional neural network model is AlexNet.

The Alex Net comprises five convolutional layers with a ReLU activation function and three fully connected layers. The Alex Net contains 62,000,000 trainable variables.

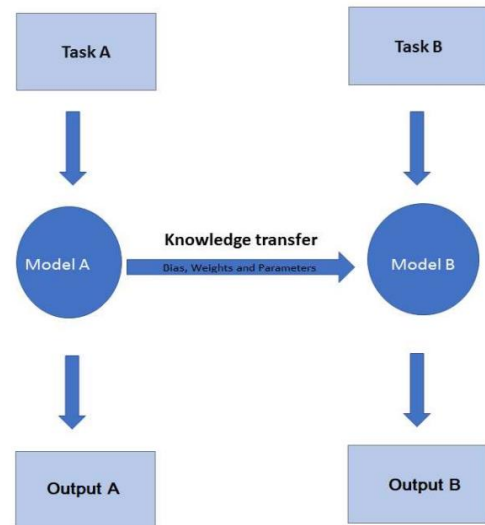


Fig. 4. Transfer learning techniques.

An example of an approach to transfer learning is shown in Fig. 4. In comparison with AlexNet, VGGNet increases performance while requiring less time to train. The convolutional and pooling layer kernels employed by VGGNet were lower than those utilized by AlexNet, indicating an important difference between the two networks. During the entire training phase, the kernel's size is fixed. VGG16 and VGG19 constitute two among the 14 distinct types of VGGNets. The number shows the number of network levels they are. There are 138 million trainable parameters in the VGG16. The ResNet addresses the vanishing gradient problem during the deep convolutional neural network training process. The ResNet makes use of a shortcut connection to improve network performance. In an entire network, there are only two pooling layers. ResNet18, ResNet50, and ResNet101 are the most common ResNet models. A total of eleven million trainable parameters in the ResNet18. In addition, the parallel kernel methods for handling flexible kernel values are presented for the inception net [10]. Google Net is the Inception Net's simplest iteration. In Google Net, there are 6.4 million trainable parameters.

DDoS volumetric attack constitutes the most harmful malicious internet traffic. This volumetric attack aims to

overwhelm the victim's computational capacity internet connections by having numerous attackers coordinate the transfer of a high rate of void data [11].

The proposed intrusion detection system for the detection of DDoS attacks and the Elliptic Curve Cryptography (ECC) - based secure encryption scheme. This method undergoes training and testing phases. In the training phase, the SDN dataset undergoes preprocessing where data nominalization, replacement of missing attributes and data normalization has been done. The proposed classification is carried out on the training phase to identify or classify the data to be normal or attacked data. Then the procedure is followed by preprocessing and classification as like in the training phase. Then if the classified data is normal, it can be further prevented from the attackers by encrypting the data using the ECC technique and stored in the network. If any need of encrypted real data in future will decrypt the data in the network and make use of it. Otherwise, the attacked (hacked) data is stored as a log file in the network for future attack detection. SDN-Data set which consists of forty-one features of is considered as normal and attacked type. Each feature is categorized into three types of attribute value types namely nominal, ordinal.

### C. Hyperparameter Optimization

It is necessary for machine learning model training to offer optimal performance. Hyperparameters are parameters that are set before training begins and control aspects of the training process that are not learned from the data. These parameters can significantly affect a model's performance, generalization, and convergence [12].

Here are the steps and techniques involved in hyperparameter optimization:

**Identify Hyperparameters:** Start by identifying the hyperparameters that need to be optimized. Depending on the type of model you are training, these could involve learning rate, batch size, number of layers, number of units in each layer, dropout rates, regularization strength, etc.

### D. Optimization Methods

**Grid Search:** For each hyperparameter, a grid of possible values needs to be specified, and all possible combinations should be thoroughly examined. It is easy to set up and can work well for a small parameter space, but it can be computationally expensive.

**Random Search:** Instead of trying all possible combinations, random search randomly selects parameter combinations to evaluate. It is more efficient than grid search for larger parameter spaces.

**Bayesian Optimization:** This is a more advanced approach that models the underlying function that maps hyperparameters to model performance. It uses this model to intelligently choose the next set of hyperparameters to evaluate, potentially reducing the number of evaluations needed.

**Gradient-Based Optimization:** Some libraries offer methods that use gradient-based optimization techniques to tune hyperparameters.

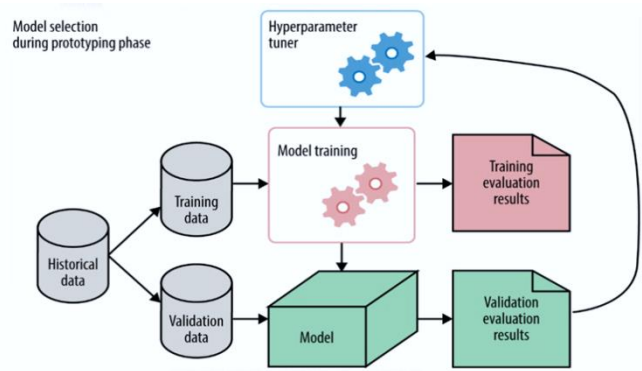


Fig. 5. Hyper parameter tuning process.

### E. Implement Hyperparameter Search

For grid search or random search, loop through different hyperparameter combinations, train the model on the training set and evaluate it on the validation set using the chosen evaluation metric. For Bayesian optimization, use a library like scikit-optimize, or Bayesian Optimization.

**Iterate and Tune:** Based on the results of the validation performance, adjust the range or values of the hyperparameters and repeat the optimization process. Fig. 5 shows the hyperparameter tuning process.

**Evaluate on Test Set:** Once you have found the best hyperparameters using the validation set, evaluate the model on a separate test set that was not used during the optimization process to get an unbiased estimate of the model's performance.

## V. EXPERIMENTAL RESULTS

After Steps to be followed

- Feature extraction.
- Data modification
- Classification
- Decision making

### A. Tools

We used i7 processor, a 256 SSD laptop TensorFlow, and Google Colab tools to train our model. We conduct perform manual attacks on DDoS and collect the data sets. We collect data sets of real-time attacks that have taken place previously from internet data sources.

### B. Preprocessing

For the purpose of minimizing noise in the data and improving the efficacy of the previously mentioned technique, preprocessing should be done on the data. Feature extraction and data transfer to numerical values are the two phases of data preparation.

### C. Feature Extraction

Feature extraction is the main step for data classification. In this project, three characteristics—packet length, delta time, and protocol—are considered in account.

D. Classification and Modification

With Data then, a numerical representation for these characteristics is created. The numerical values are fed to the ML models, with 30% utilized as training data and 70% for model testing. Each ML model is processed concurrently.

E. Decision Making

The voting majority among all algorithms determines the result.

Fig. 6 shows the training of the hyper model and Fig. 7, Fig. 8 shows the graph drawn between accuracy of the hyper model and epochs and loss vs. Epoch graphs.

```

Epoch 89/100
2272/2272 - 5s - loss: 0.0180 - accuracy: 0.9928 - val_loss: 0.0197 - val_accuracy: 0.9921 - 5s/epoch - 2ms/step
Epoch 90/100
2272/2272 - 5s - loss: 0.0218 - accuracy: 0.9916 - val_loss: 0.0206 - val_accuracy: 0.9917 - 5s/epoch - 2ms/step
Epoch 91/100
2272/2272 - 6s - loss: 0.0200 - accuracy: 0.9916 - val_loss: 0.0204 - val_accuracy: 0.9914 - 6s/epoch - 3ms/step
Epoch 92/100
2272/2272 - 4s - loss: 0.0202 - accuracy: 0.9917 - val_loss: 0.0235 - val_accuracy: 0.9897 - 4s/epoch - 2ms/step
Epoch 93/100
2272/2272 - 4s - loss: 0.0198 - accuracy: 0.9920 - val_loss: 0.0232 - val_accuracy: 0.9910 - 4s/epoch - 2ms/step
Epoch 94/100
2272/2272 - 6s - loss: 0.0191 - accuracy: 0.9923 - val_loss: 0.0240 - val_accuracy: 0.9914 - 6s/epoch - 3ms/step
Epoch 95/100
2272/2272 - 4s - loss: 0.0187 - accuracy: 0.9920 - val_loss: 0.0215 - val_accuracy: 0.9915 - 4s/epoch - 2ms/step
Epoch 96/100
2272/2272 - 6s - loss: 0.0204 - accuracy: 0.9922 - val_loss: 0.0231 - val_accuracy: 0.9908 - 6s/epoch - 3ms/step
Epoch 97/100
2272/2272 - 5s - loss: 0.0182 - accuracy: 0.9926 - val_loss: 0.0258 - val_accuracy: 0.9910 - 5s/epoch - 2ms/step
Epoch 98/100
2272/2272 - 4s - loss: 0.0186 - accuracy: 0.9924 - val_loss: 0.0252 - val_accuracy: 0.9904 - 4s/epoch - 2ms/step
Epoch 99/100
2272/2272 - 6s - loss: 0.0183 - accuracy: 0.9925 - val_loss: 0.0223 - val_accuracy: 0.9916 - 6s/epoch - 3ms/step
Epoch 100/100
2272/2272 - 4s - loss: 0.0201 - accuracy: 0.9921 - val_loss: 0.0313 - val_accuracy: 0.9907 - 4s/epoch - 2ms/step
    
```

Fig. 6. Training of proposed model.

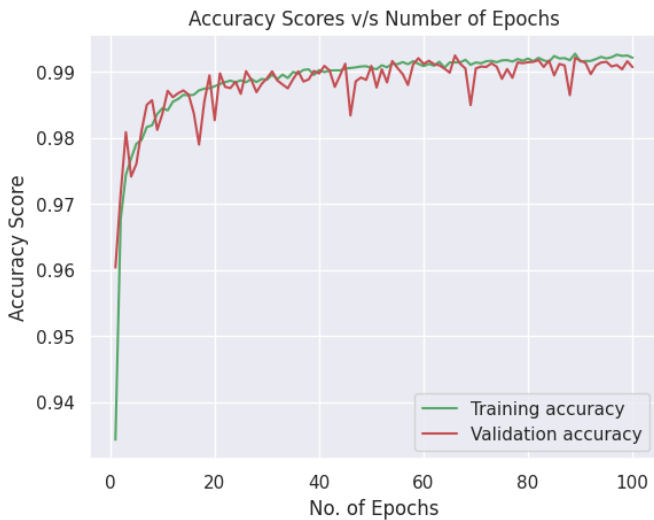


Fig. 7. Accuracy v/s Epochs in hyper model.

F. AUC and ROC

Evaluation of performance is an essential function in machine learning as shown in Fig. 9. So, we can depend on an AUC-ROC Curve when it comes to the classification their task. The AUC (Area Under the Curve) and ROC (Receiver Operating Characteristics) curves is employed to evaluate or show the performance of the multi-class classification trouble.

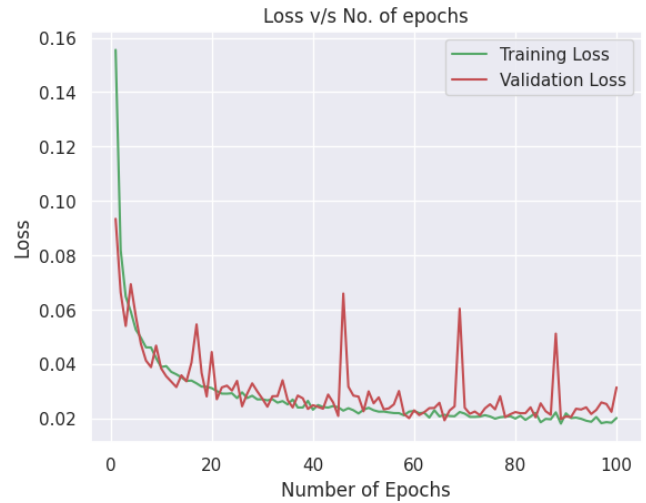


Fig. 8. Loss Vs Epochs.

It is one of the most essential criteria for assessing the efficiency of any classification model. AUROC (Area Under the Receiver Operating Characteristics) is a different method of expressing its contents.

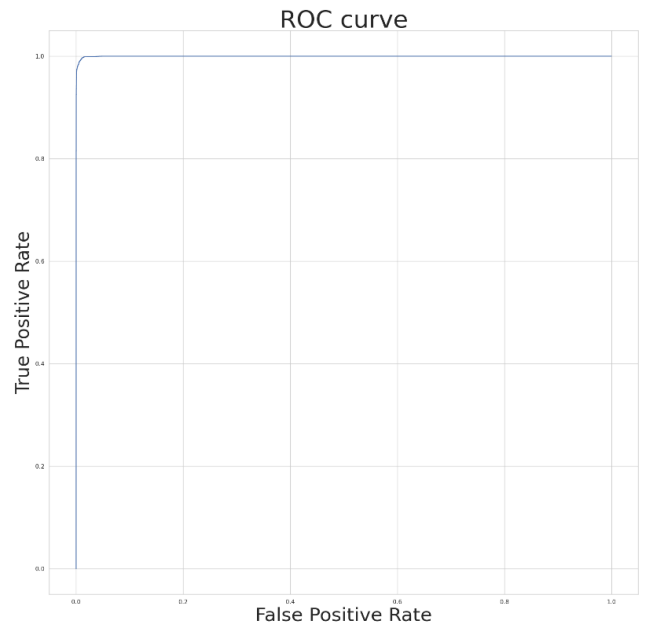


Fig. 9. ROC curve of Hyper model.

The accuracy of the proposed algorithm is compared with other standard machine learning algorithms in Table I.

TABLE I. COMPARISON OF ACCURACY OF DIFFERENT MODEL

S.no	Model	Accuracy
1	XGBoost	98.179892
2	KNN	96.809194
3	Decision Tree	96.619800
4	Proposed Hypermodel	99.072289

## VI. CONCLUSION

Nearly all the work needs to be done manually while utilizing traditional methods, and the results are often inaccurate and difficult to identify risks. However, by employing a Machine Learning strategy, we can spot risks much more quickly than with traditional methods. Three neural network (ML) algorithms—Naive Bayesian, KNN, and Random Forest were employed in our proposal. The suggested hypermodel produced highly excellent precision and other output metrics like AUC and ROC by utilizing the hyper-tuning method in machine learning. We will be mounting it on all networks such routers, firewalls, and server since it has a limited resource consumption and can run on inexpensive components. The current system simply detects DDoS attacks, but in the future, we will improve it to a DDoS-avoiding model to improve server security. As technology advances, we will boost this model's efficiency.

## REFERENCES

- [1] G. Asad M, Asim M, Javed T, Beg MO, Mujtaba H, Abbas S (2020) Deep Detect: detection of Distributed Denial of Service attacks using deep learning. *Computer J* 63:983–994
- [2] V. S, B. A. N.G. and N. -K. Baik, "Detection of DoS Attacks in Smart City Networks with Feature Distance Maps: A Statistical Approach," in *IEEE Internet of Things Journal*, doi: 10.1109/IJOT.2023.3264670.
- [3] Mishra, A.; Gupta, N.; Gupta, B.B. Defensive mechanism against DDoS attack based on feature selection and multi-classifier algorithms. *Telecommun. Syst.* 2023, 82, 229–244
- [4] A. Haddaji, S. Ayed and L. C. Fourati, "A Transfer Learning Based Intrusion Detection System for Internet of Vehicles," 2023 15th International Conference on Developments in eSystems Engineering (DeSE), Baghdad & Anbar, Iraq, 2023, pp. 533-539, doi: 10.1109/DeSE58274.2023.10099623.
- [5] Farhan Ullah, Shamsher Ullah, Gautam Srivastava, Jerry Chun-Wei Lin, IDS-INT: Intrusion detection system using transformer-based transfer learning for imbalanced network traffic, *Digital Communications and Networks*, 2023,ISSN 2352-8648, <https://doi.org/10.1016/j.dcan.2023.03.008>.
- [6] Bhardwaj A, Mangat V, Vig R (2020) Hyperband tuned deep neural network with well posed stacked sparse AutoEncoder for detection of DDoS attacks in Cloud. *IEEE Access* 8:181916–181929
- [7] Catak FO, Mustacoglu AF (2019) Distributed denial of service attack detection using autoencoder and deep neural networks. *J Intell Fuzzy Syst* 37:3969–3979
- [8] Chen J, tao Yang Y, ke Hu K, bin Zheng H, Wang Z (2019) DADMCNN: DDoS attack detection via multi-channel CNN. In: *ACM international conference proceeding series, vol Part F1481*. Association for Computing Machinery, New York, pp 484–488
- [9] Morgan Kaufmann, pp 1–38 Hasan MZ, Hasan KMZ, Sattar A (2018) Burst header packet flood detection in optical burst switching network using deep learning model. *Procedia Comput Sci* 143:970–977
- [10] He J, Tan Y, Guo W, Xian M (2020) A small sample DDoS attack detection method based on deep transfer learning. In: *Proceedings—2020 International Conference on Computer Communication and Network Security, CCNS 2020*. Institute of Electrical and Electronics Engineers Inc., pp 47–50
- [11] Nugraha B, Murthy RN (2020) Deep learning-based slow DDoS attack detection in SDN-based networks. In: *2020 IEEE conference on Network Function Virtualization and Software Defined Networks, NFV-SDN 2020—Proceedings*. Institute of Electrical and Electronics Engineers Inc., pp 51–56
- [12] Panigrahi R, Panigrahi R, Borah S (2018) A detailed analysis of CICIDS2017 dataset for designing intrusion detection systems. *Int J Eng Technol* 7:479–482 Premkumar M, Sundararajan TV (2020) DLDM: deep learning-based defense mechanism for denial-of-service attacks in wireless sensor networks. *Microprocess Microsyst* 79:103278

# Cyberbullying Detection Based on Hybrid Ensemble Method using Deep Learning Technique in Bangla Dataset

Md. Tofael Ahmed<sup>1</sup>, Afroza Sharmin Urmi<sup>2</sup>, Maqsurur Rahman<sup>3</sup>, Abu Zafor Muhammad Touhidul Islam<sup>4</sup>,  
Dipankar Das<sup>5</sup>, Md. Golam Rashed<sup>6</sup>

Department of Information and Communication Technology, Comilla University, Bangladesh<sup>1, 2, 3</sup>

Department of Electrical & Electronics Engineering, University of Rajshahi, Bangladesh<sup>4</sup>

Department of Information and Communication Engineering, University of Rajshahi, Bangladesh<sup>1, 5, 6</sup>

**Abstract**—Globalization is certainly a blessing for us. Still, this term also brought such things that are constantly not only creating social insecurities but also diminishing our mental health, and one of them is Cyberbullying. Cyberbullying is not only a misuse of technology but also encourages social harassment among people. Research on Cyberbullying detection has gained increasing attention nowadays in many languages, including Bengali. However, the amount of work on the Bengali language compared to others is insignificant. Here we introduce a Hybrid ensemble method using a voting classifier in Bangla Cyberbullying detection and compare this with traditional Machine Learning and Deep Learning Classifiers. Before implementation, Exploratory Data Analysis was performed on the dataset to gather better insight. There are lots of papers that have already been published in other languages where it is seen that the hybrid approach provides better outcomes compared to traditional methods. Thus, we propose a highly well-driven method for Cyberbullying detection on the Bangla dataset using the hybrid ensemble method by voting classifier. The overall deployment consists of three Machine Learning classifiers, three Deep Learning classifiers, and a Hybrid approach using the voting classifier. Finally, the Hybrid ensemble method yields the best performance with an accuracy of 85%, compared with other Machine and Deep Learning methods.

**Keywords**—Bangla dataset; cyberbullying; exploratory data analysis; machine learning; deep learning; hybrid ensemble method

## I. INTRODUCTION

Globalization impacted Bangladesh immensely a long time ago, and we're moving forward with this new era and keeping ourselves up-to-date with technology, which is a blessing and positive news for us. But now the major concern is cyberbullying, which is also spreading drastically in the Bengali language. People got so aggressive while bullying others, and it is high time we should be concerned and do more research [1] on our language to prevent and detect those nonsocial texts. Research says, this type of behavior recurrently arose on YouTube, Facebook, and Twitter sites (Eric Rice, Cyberbullying perpetration and victimization among middle-school students., 2015). There are about 126.21 million internet consumers in Bangladesh [2], and the majority of active internet consumers are young. Also, Bengali is in the 6th rank as the most spoken language around the globe [3],

which accelerates the use of Bangla over social media. Which also increases the devastating amount of bullying on the internet. The expression "cyberbullying" can refer to a variety of behaviors, including hostile material, harassment, toxic commenting (such as gibe, triggered, sexual, or religious), and so on. And these sorts of aggressive behavior most lead to terrible mental health issues, such as self-harm anxiety, depression, social and emotional perplexity even suicidal thoughts or attempted suicide [4]. In Bangladesh, a survey stated, about 85 percent of youths believe that online bullying is an unadorned problem and 8 percent of youths have faced online bullying, at least one time a week or more since the pandemic [5].

Therefore, to mitigate such heinous acts of cyberbullying, many global preventive and intervention approaches have been introduced to improve the safety of internet users all around the world, and we should also be concerned about our Bangla language. Thus, we are approaching a new method to prevent this vile activity. Due to the numerous benefits the hybrid approach has over traditional machine learning algorithms, researchers are moving away from traditional machine learning techniques and toward them in the detection of cyberbullying. That's why, focusing on detecting bully on Bengali Language using a Hybrid Ensemble approach. The following are main highlights of this paper:

1) Introducing and proposing a hybrid ensemble approach that combines all the traditional approaches utilized in this study and uses a voting classifier to detect cyberbullying in the Bangla dataset.

2) The proposed method outperforms the currently used worldwide for classification and analysis, including Dense Architectural, LSTM, Bi-LSTM, KNN, Logistic Regression, and Decision Tree.

3) Performed exploratory data analysis (EDA) to get the appropriate insights and visuals for the desired cyberbullying model in the Bangla dataset.

4) Performed visualization and comparative analysis on the classification performance of three traditional machine learning and deep learning algorithms.

5) Evaluated various feature extraction methods to identify the optimal strategies for feature extraction and text

embedding for both conventional machine learning and neural network-based techniques.

Our whole work is divided into two parts:

- Performing Exploratory Data Analysis (EDA) to get insight from the dataset.
- And then introducing a hybrid ensemble method to detect cyberbullying in the Bangla dataset.

## II. RELATED WORKS

Cyberbullying is indeed like a parasite in our modern tech world. There are so many works and activities being practiced preventing this. Recently, the amount of work has been increased in the field of Cyberbullying detection in the Bangla dataset. But the progress is not comprehensive; we should enrich our research by doing more work to prevent this aggressive vulnerability on social media.

For English language, tons of work already been done in text categorization or Cyberbullying detection using text mining by classifying posts or conversations. Yin et al. text classification by using supervised learning [6]. In [7], they proposed unsupervised method for bully detection. Haidar et al. detect Cyberbullying by proposing a multilingual system in the Arabic language by using Machine Learning [8]. Dadvar et al. they proposed a Hybrid approach to detect Cyberbullying, where they show that the hybrid approach provides better performance than the expert systems [9]. Zhao et al. proposed an automatic Cyberbullying detection using bullying features on social media [10]. Mohammed Ali Al-garadi et al. proposed a model that provides a practical explanation for Cyberbullying detection in social media. This is an offensive language detection approach equipped with a lexical syntactic feature only [11]. Dalvi et al. they introduced a method to recognize and stop Cyberbullying on Twitter using machine learning [12]. Muneer et al. proposed an automatic bully detection system where they used more than 35,000 distinctive tweets as a dataset [13]. Herath et al. presented an automated bully detection system against immigrants, women, and cross-domain adaptability [14]. Robin M. Kowalski et al. came out social media bullying between Middle School Students, their results disclosed that about 84% of school students have experienced bullying in this study [15]. In a previous study conducted by Prathyusha et al. [16], a novel approach was proposed that integrates the Multiple Correlation Coefficient and the Support Vector Machine. In Bengali language, some noteworthy works been seen for Cyberbullying detection. Mahmud et al. [17] Utilizing the Bangla dataset, we were able to come up with a visualization method work on the Bangla corpus is specifically used for sentiment analysis [18]. In [19], Shahin Akhter et al. performed Cross-validation using ML classification models with 2400 data labeled as bullied and non-bullied and achieves superior performance on Bangla text with a detection accuracy of 97%. This study shows a very insignificant amount of data, that's why it's over fitted. In [20], Ahmed et al. used 44,001 users' comments from popular public Facebook pages and came out with a binary classification model of 87.91% accuracy. They proposed binary classification model which isn't well suited for text classification. In [21], Chakraborty et al. used several ML and

DL approach to analyze Bangla texts. In [22], Ahammed et al. proposed a Machine Learning approach with an accuracy of 72%. Again in [23] [24], Ahmed et al. used Bangla, Romanized Bangla and Meme Detection [25] text for Cyberbullying detection using traditional ML, and DL Classifiers.

In terms of bullying detection, the above approaches are all good, but they got a few lacking like, using insignificant amount of dataset, implementing traditional approach for bully detection, and the approach that doesn't help to classify text accordingly. Hence, we're introducing a new approach for Cyberbullying detection on the Bangla dataset: Hybrid Ensemble Method using the voting classifier. The detailed methodology is discussed in Section III. Section IV presents the Results analysis and the conclusion is given in Section V.

## III. RESEARCH METHODOLOGY

This research mainly focuses on introducing a new approach for Cyberbullying detection in the Bangla dataset.

### A. Dataset

There is total number of 10,512 data on our dataset and the dataset consists of three columns - comments, class, and gender. The class column categorized into- Gibe, Triggered, Sexual, Religious, and not bully. And the Class and Gender section is specifically used to perform Exploratory Data Analysis to get insight from the Bangla dataset. Fig. 1, a complete portrait of how the bully is differing from gender to gender.

Gender	Class				
	Gibe	Not bully	Religious	Sexual	Triggered
Female	1,321	1,101	286	251	873
Male	1,824	3,174	439	242	1,021

Fig. 1. Different types of bullies among gender.

### B. Model Procedure

In the proposed work, introducing Hybrid Ensemble approach for bully detection in the Bangla dataset. The overall reasearch is divided into four parts or sections:

- Exploratory Data Analysis to get proper insight from the dataset,
- Machine Learning deployment,
- Deep Learning deployment, and
- Hybrid Ensemble approach deployment.

To get an accurate and precise model, one need to know the dataset precisely that's why exploratory data analysis is must to do part before any model implementation. In our work, we use the table for our data visualization and find out proper insight through this visualization. The Bangla comments are collected from social media platform. Since



dataset is in raw format, contain noisy elements, unwanted data, null value which could degrade our model evaluation. So, lots of data preprocessing method been applied to sort out the dataset and remove those unwanted part from the Bangla dataset. A classifier is a function that uses an example's values as predictor variables or independent variables to determine the class to which the example belongs (the dependent variable). A computational software called machine learning is able to learn without being told where to look. In our work, Machine Learning Classifier perform one part which helps to understand further classification. New and robust deep learning (DL) algorithms have been developed as a result of advancements in computing technology, and they have shown promising outcomes in a variety of applications. Further implemented a few Deep Learning model in order to get overview from the dataset.

Finally introduced a Hybrid Ensemble method and compare between these three methods, to find out best performance. Illustrating overall workflow in below Fig. 2. Based on split data, the program then trains each classifier to develop a method of classifying. Three columns from dataset is taken for further analysis. They are- comments, Gender, and class. The class column divided comments into five categories:

Comments can be in different formats, like:

- Not bully,
- Triggered,
- Gibe,
- Sexual,
- Religious, etc.

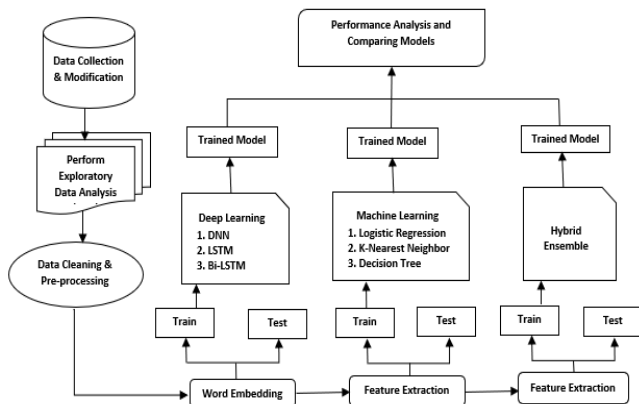


Fig. 2. Diagram of the proposed workflow for the research.

So, Not Bully are the type of comments which are in general comments on a post or write as well wishing. Then others four types of comments are basic criteria of bully comment or language. For better clarification triggered comments are like make some pointing negatively publicly, gibe means spreading some rumors about other user, sexual mean abuse someone or eve teasing female through social media and the religious bullies are like contradicting with religious and conflict with each other and so on.

For further implementation and evaluation, we used three ML classifiers and three DL classifiers and Hybrid Ensemble approach using voting classifier. The accuracy, precision, recall, f1 score, validation loss, and accuracy of each of the classifiers used in these models have been evaluated in order to choose the optimal model for implementing the Cyberbullying detection on the Bangla dataset. The model procedure is detailed below.

1) *Exploratory data analysis:* The Exploratory Data Analysis (EDA) is performed on Tableau and came out with some important points and insights from our dataset. Dashboard of the overall Exploratory Data Analysis is shown in Fig. 3. Which is a collection or assembly of multiple worksheets. Worksheet are made by importing Class and Gender columns in Tableau.

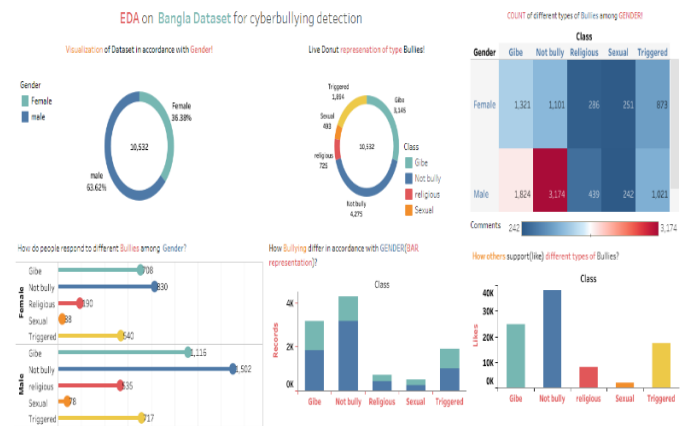


Fig. 3. Exploratory data analysis (EDA) on the Bangla dataset, and dashboard using tableau.

2) *Data analyzing and preprocessing:* For analysis, first dropped all the unwanted columns from the dataset. Before implementation, data analysis is mandatory for getting a clean and accurate model. The BNL corpus is used to remove Bangla stop-words, letters, and digits from the Bangla comments. The dataset was also cleaned of any duplicate information, links, or URLs, as well as numerals, punctuation marks, and emojis. The next step is to clean, instance selection, remove noisy data, unwanted value, feature extraction, and selection are among the steps that are involved.

3) *Feature extraction:* Machine Learning and Hybrid Ensemble approach feature extraction, we used Term Frequency & Inverse Document. It is stated that, in Machine Learning strings is converted into number by using TF-IDF and then provide a numerical format for the Machine Learning models. Using the text vectorizer term frequency-inverse document frequency, text is transformed into a vector format. The concepts of Term Frequency (TF) and Document Frequency are integrated (IDF).

4) *Split processed dataset:* The dataset was divided into training and testing data using the sklearn model selection module for subsequent implementation. A total number of 10,512 data is in the dataset and split them for further training and testing using several approaches.

### C. Machine Learning Classifiers

Here three Machine Learning classification models been used for the bully detection purpose and they are: Logistic Regression (LR), K-Nearest Neighbor (KNN), and Decision Tree (DT).

1) *Logistic regression (LR)*: Logistic Regression (LR) been widely used for classification algorithm (two-class classification). Logistic Regression estimates the probability of an event occurring by using the sigmoid function which maps any real value into another value between probability 0 and 1. The sigmoid function is written in Eq. (1).

$$\text{Logit}(x) = 1/(1 + \exp(-x)) \quad (1)$$

2) *K-Nearest neighbor (KNN)*: The k-nearest neighbors' algorithm (KNN) is a supervised learning and a non-parametric classifier, which uses propinquity where they group distinct data point to make prediction. Although it can be applied to classification or regression problems, it is commonly employed as a classification algorithm because it assumes that comparable points are located close to one another.

3) *Decision tree (DT)*: Decision Tree is also a supervised learning classifier. When it comes to classification issues, decision tree is often used. In DT, internal nodes outline dataset appearances, branches signify decision rules, and each leaf node denotes the result.

### D. Deep Learning Classifiers

For our research purpose, we used three Deep Learning approaches, like- Dense architecture model, LSTM, and Bi-LSTM.

1) *Dense architecture model*: The proposed approaches run 30 epochs to find out the upmost accurate performance for Dense Architectural model. It can be observed that, by the time each cycle or epoch running the performance of the model also enhances, which means decreasing validation loss of model and increasing validation accuracy of the model. The model summary provides the layer, shape, and number of parameters used in each layer. Fig. 4 below demonstrating, overall Dense Architectural Model summary, which includes that the model is running sequentially, the embedding layer got total 8000 parameters, dense layer got 408 parameters, there is no dropout in the dense layer. It also indicates that the total parameters are 8433 and all of them are trainable.

2) *Long short-term memory (LSTM)*: Long short-term memory (LSTM) is a type of recurrent neural network that was created specifically to stop the neural network output for a particular input from expanding up as it rotates with the feedback loops. Recurrent networks were able to outperform other neural networks at pattern identification thanks to these feedback loops. Another popular advanced recurrent neural network (RNN) structure that was created with long-range dependencies and temporal sequences in mind is known as Long Short-Term Memory. Fig. 5, Number of epochs and validation accuracy and loss scores in our LSTM model.

Model: "sequential"

Layer (type)	Output Shape	Param #
embedding (Embedding)	(None, 50, 16)	8000
global_average_pooling1d (G1)	(None, 16)	0
dense (Dense)	(None, 24)	408
dropout (Dropout)	(None, 24)	0
dense_1 (Dense)	(None, 1)	25
Total params: 8,433		
Trainable params: 8,433		
Non-trainable params: 0		

Fig. 4. Proposed dense architectural model.

```
Epoch 1/30
263/263 - 37s - loss: 0.6382 - accuracy: 0.6522 - val_loss: 0.5797 - val_accuracy: 0.7291
Epoch 2/30
263/263 - 16s - loss: 0.5639 - accuracy: 0.7448 - val_loss: 0.5576 - val_accuracy: 0.7457
Epoch 3/30
263/263 - 13s - loss: 0.5613 - accuracy: 0.7405 - val_loss: 0.5762 - val_accuracy: 0.7316
Epoch 4/30
263/263 - 14s - loss: 0.5440 - accuracy: 0.7563 - val_loss: 0.5551 - val_accuracy: 0.7489
Epoch 5/30
263/263 - 13s - loss: 0.5212 - accuracy: 0.7702 - val_loss: 0.5408 - val_accuracy: 0.7468
Epoch 6/30
```

Fig. 5. Validation accuracy and loss scores of LSTM model, and number of epochs.

3) *Bi-LSTM*: Bidirectional recurrent neural networks are the combination of two independent RNNs together for further implementation. With the help of this Bi-LSTM structure, the networks can vividly have both forward and backward information at every stage of the process. Using a bidirectional will run inputs in two ways: one from the past to future and another one from the future to past and where LSTM that only runs backward (unidirectional). Fig. 6 illustrates the number of epochs and validation accuracy and loss scores in our Bi-LSTM model.

```
Epoch 1/30
263/263 - 43s - loss: 0.6297 - accuracy: 0.6539 - val_loss: 0.5504 - val_accuracy: 0.7380
Epoch 2/30
263/263 - 22s - loss: 0.5224 - accuracy: 0.7536 - val_loss: 0.5296 - val_accuracy: 0.7456
Epoch 3/30
263/263 - 19s - loss: 0.5074 - accuracy: 0.7576 - val_loss: 0.5211 - val_accuracy: 0.7446
Epoch 4/30
263/263 - 18s - loss: 0.4950 - accuracy: 0.7658 - val_loss: 0.5271 - val_accuracy: 0.7476
Epoch 5/30
263/263 - 17s - loss: 0.5002 - accuracy: 0.7636 - val_loss: 0.5927 - val_accuracy: 0.6981
```

Fig. 6. Validation of the correctness of the Bi-LSTM model in terms of loss as well as the number of epochs.

### E. Hybrid Ensemble Method

Hybrid Ensemble Method is one of the popular methods used nowadays for better predicted outcome. It's also called the heterogeneous assembly of weak learners. Lots of Machine Learning classifiers are combined in this task to come up with a classification problem. Ensemble learning techniques got a long record and history of showing handsome performance comparing with any other traditional

ML approaches. The domains of these applications include classification and regression problems. Fig. 7 shows how the hybrid ensemble method works in a simple way.

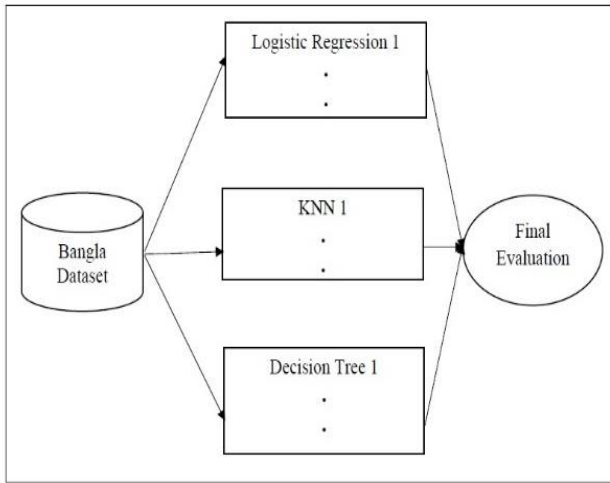


Fig. 7. Basic working procedure of hybrid ensemble method.

#### IV. RESULT ANALYSIS

##### A. ML Models Result Analysis

The performance of the Machine Learning models was further examined through the utilisation of Receiver Operating Characteristic (ROC) curves and the Area Under the ROC Curve (AUROC). In Fig. 8, the receiver operating characteristic (ROC) curve and the area under the ROC curve (AUROC) are depicted for our machine learning classifiers. It is observed that the logistic regression (LR) model achieved the highest AUROC score of 88%. It has been noticed that Logistic Regression achieved the greatest AUROC score. It's stated that, highest AUROC means better classifier. Summary of ROC curve with AUROC score: Logistic Regression AUROC score is 0.88, K-Nearest Neighbor AUROC score is 0.832, and Decision Tree score is 0.551. According to the measurements of AUROC, Logistic Regression is the best classifier amongst the three ML classifiers.

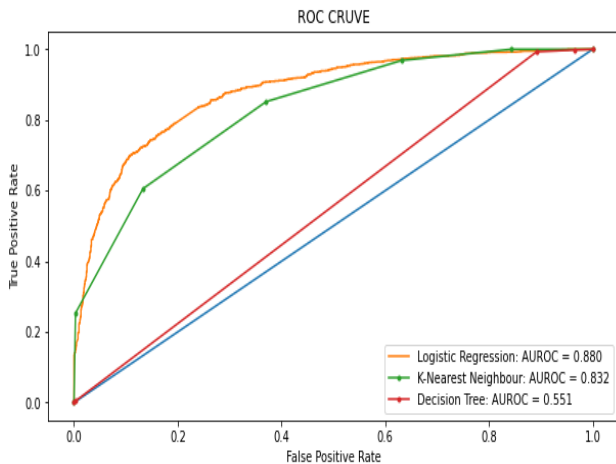


Fig. 8. Receiver operating characteristic, and area under-ROC curve and scores of ML classifiers.

##### B. DL Models Result Analysis

For Deep Learning models, the visualization shown below is the result in validation loss and accuracy curve for comparison. Both Validation loss and Validation accuracy of Dense Architectural Model are illustrating in Fig. 9. For Dense Architectural Model with increasing number of epochs validation loss also decreasing and in mean-time validation accuracy increased. For the Dense Architectural Models Training and Validation evaluation, where employed a total number of 12 epochs. In below graph, validation accuracy versus the number of epochs provides a visual. Where blue line represents as Training accuracy score and yellow line representing Validation Accuracy score. This is time series analysis of visualizing how accuracy scores of Dense Architecture model increasing with number of epochs. Highest accuracy for both training and validation is highlighting in epoch number 12, where Training accuracy score around 0.775 and Validation accuracy score is around 0.770.

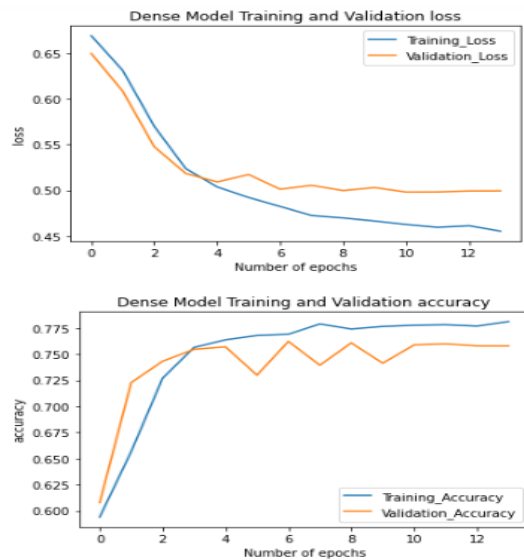


Fig. 9. Time series representation Dense Architectural model (Validation loss and accuracy scores with respect to number of epochs).

LSTM model Validation loss and accuracy in below Fig. 10 Validation accuracy as a function of the total number of epochs. Where blue line represents as Training accuracy score and yellow line represents Validation Accuracy score. This is a time series analysis of visualizing how accuracy scores of LSTMs (Long Short-Term Memory) model increasing with number of epochs. Highest accuracy for both training and validation is highlighted in epoch number 5, where Training accuracy score around 0.78 and Validation accuracy score is around 0.76 in the 4th epoch.

Bi-LSTM model Validation loss and accuracy in the Fig. 11. Where blue line represents as Training accuracy score and yellow line represents Validation Accuracy score. This is a time series analysis of visualizing how accuracy scores of Bi-LSTM (Bidirectional) model increase with number of epochs. Highest accuracy for both training and testing is highlighted in epoch number 7, where Training accuracy score around 0.78 and Validation accuracy score is around 0.7 at the 7th epoch.

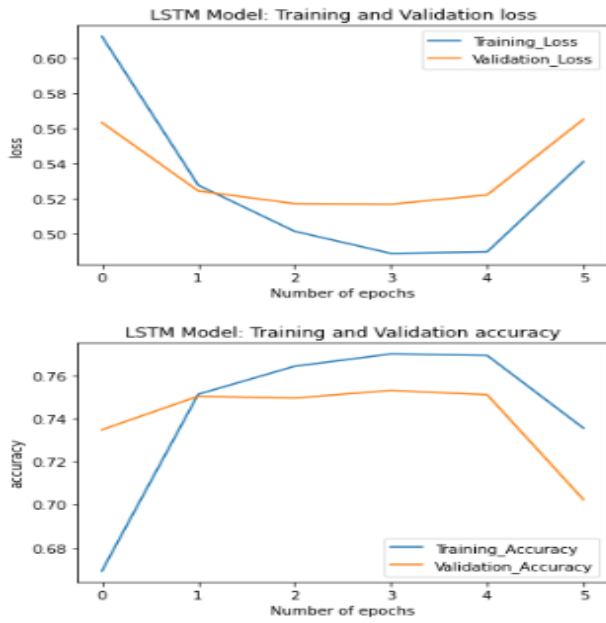


Fig. 10. Time series chart of LSTM model (Validation loss and accuracy scores with respect to number of epochs).

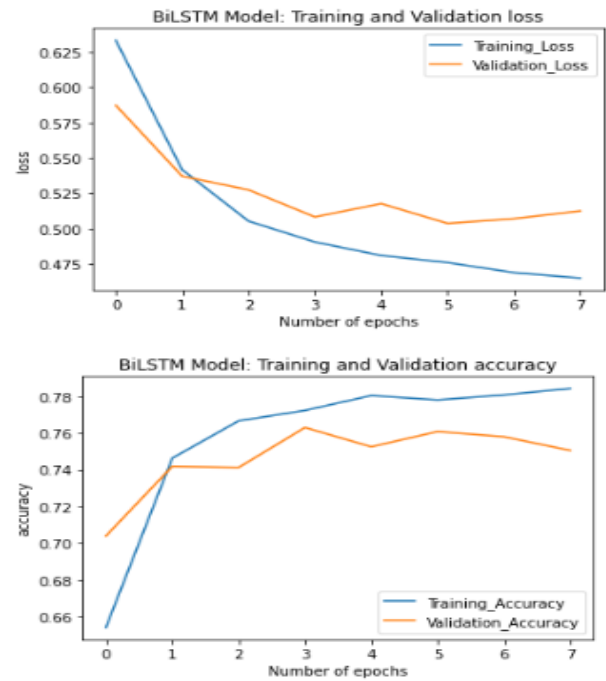


Fig. 11. Time series representation Bi-LSTM model (Validation loss and accuracy scores with respect to number of epochs).

C. Hybrid Ensemble Models Result Analysis

Now finally comparing and analyzing the proposed Hybrid ensemble model with ML and DL Models result. In Fig. 12, a bar chart representation to compare performance measurement scores of Machine learning classifiers and the proposed Hybrid Ensemble model is shown. Where it reflects that Hybrid Ensemble model performs better than Machine Learning classifier.

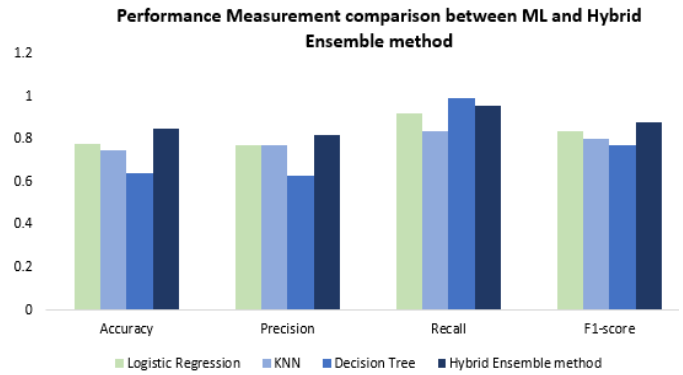


Fig. 12. Comparing our proposed hybrid ensemble approach with machine learning models.

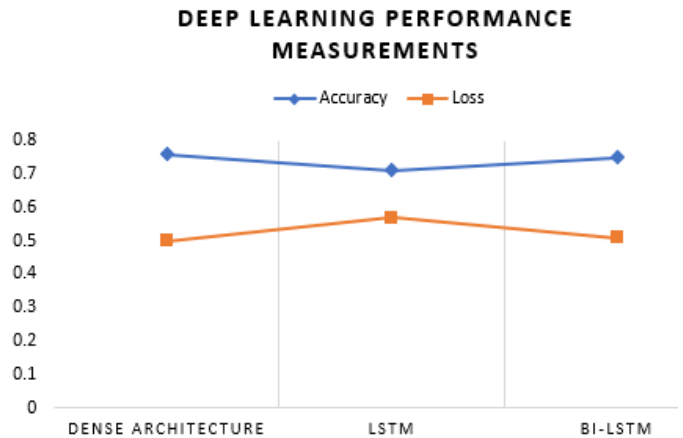


Fig. 13. Time series chart of deep learning models.

Fig. 13 illustrates a short time series chart comparing Deep Learning models with our approached Hybrid Ensemble method. Where blue line representing accuracy scores and yellow line representing loss performance scores of the deep learning models.

V. CONCLUSION

The primary objective of this study is to priorities efforts towards the identification of Cyberbullying using the Bangla dataset, while also emphasizing the importance of raising awareness using visualization techniques. This research presents a novel methodology for identifying instances of bullying in the Bengali language. The Hybrid Ensemble approach was presented, which use a voting classifier and achieves an accuracy rate of 85%. In addition, conventional machine learning classifiers and deep learning classifiers were utilized for comparison with the proposed methodology. Ultimately, after doing a comparative analysis of the three ways, it has been determined that the Hybrid Ensemble approach exhibits superior performance. The primary objective and approach of this study are to foster increased participation among researchers in their respective native languages and mitigate societal obstacles such as Cyberbullying. In subsequent investigations, there is potential for the expansion of research through the utilization of Unsupervised or Reinforcement Learning methodologies. This

approach might be employed to further boost the efficacy of the bully detection model by using a Bangla dataset.

#### REFERENCES

- [1] R. P. H. R. H. W. J. G. Eric Rice, "Cyber bullying perpetration and victimization among middle-school students.," American Journal of Public Health (ajph), pp. e66-e72, 2015.
- [2] BTRC, "http://www.btrc.gov.bd/site/page/347df7fe-409f-451e-a415-65b109a207f5/-," BTRC, Bangladesh, 2022.
- [3] edudwar.com, "https://www.edudwar.com/list-of-most-spoken-languages-in-the-world/," edudwar, 2022.
- [4] S. B. P. R. K. P. Kazi Saeed Alam, "Cyberbullying Detection: An Ensemble Based Machine Learning Approach," in 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), Tirunelveli, India, 2021.
- [5] daily-sun.com, "Online bullying a serious problem in Bangladesh: Telenor Survey," daily-sun, Bangladesh, 2021.
- [6] Z. X. L. H. B. D. D. A. K. a. L. E. Dawei Yin, "Detection of Harassment on Web 2.0," in Content Analysis in the WEB 2.0 (CAW2.0) Workshop, Madrid, Spain, 2009.
- [7] S. Immanuvelraj Kumar, "Unsupervised Hybrid approaches for Cyberbullying Detection in Instagram General Terms Cyberbullying Detection in Instagram," International Journal of Computer Applications, vol. 174, no. 26, pp. 40-46, 2021.
- [8] C. M. A. S. Batoul Haidar, "A Multilingual System for Cyberbullying Detection: Arabic Content Detection using Machine Learning," Advances in Science Technology and Engineering Systems Journal, vol. 2, no. 6, pp. 275-284, 2017.
- [9] R. B. T. F. M. d. J. M. Dadvar, "Experts and Machines against Bullies: A Hybrid Approach to Detect Cyberbullies," in Advances in Artificial Intelligence. Canadian AI, Canada, 2014.
- [10] A. Z. K. M. Rui Zhao, "Automatic Detection of Cyberbullying on Social Networks based on Bullying Features," in Proceedings of the 17th International Conference on Distributed Computing and Networking, 2016.
- [11] M. A. A.-g. a. K. D. V. a. S. D. Ravana, "Cybercrime detection in online communications: The experimental case of cyberbullying detection in the Twitter network," Comput. Hum. Behav. Algaradi2016CybercrimeDI, vol. 63, pp. 433-443, 2016.
- [12] S. B. C. A. H. Rahul Ramesh Dalvi, "Detecting A Twitter Cyberbullying Using Machine Learning," in 4th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2020.
- [13] S. M. F. Amgad Muneer, "A Comparative Analysis of Machine Learning Techniques for Cyberbullying Detection on Twitter," Future Internet, vol. 12, p. 187, 2020.
- [14] T. A. H. D. C. T. Mahen Herath, "Automated Detection of Cyberbullying Against Women and Immigrants and Cross-domain Adaptability," in Proceedings of the The 18th Annual Workshop of the Australasian Language Technology Association, Virtual Workshop, 2020.
- [15] S. P. L. P. W. A. Robin M. Kowalski, "Cyberbullying: Bullying in the Digital Age," Journal of Blackwell Publishing, vol. 7, no. 2, pp. 9-17, 2019.
- [16] R. S. I. Tata Prathyusha, "Cyberbully Detection Using Hybrid Techniques," Research gate, 2018.
- [17] Z. M. Z. T. B. A. A. N. R. A. H. F. B. A. Md Faisal Ahmed, "Bangla Text Dataset and Exploratory Analysis for Online Harassment Detection," Research Gate, 2021.
- [18] H. K. Z. H. M. B. S. M. A. I. A. I. Fuad Rahman, "An Annotated Bangla Sentiment Analysis Corpus," in International Conference on Bangla Speech and Language Processing (ICBSLP), Sylhet, Bangladesh, 2019.
- [19] S. A. Abdhullah-Al-Mamun, "Social media bullying detection using machine learning on Bangla text," in 10th International Conference on Electrical and Computer Engineering (ICECE), Dhaka, Bangladesh, 2018.
- [20] Z. M. Z. T. B. A. A. N. R. A. H. F. B. A. Md Faisal Ahmed, "Cyberbullying Detection Using Deep Neural Network from Social Media Comments in Bangla Language," Research gate, 2018.
- [21] M. H. S. Puja Chakraborty, "Threat and Abusive Language Detection on Social Media in Bengali Language," in 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), Dhaka, Bangladesh, 2019.
- [22] S. R. M. N. M. C. S. Ahammed, "Implementation of machine learning to detect hate speech in bangla language," in 8th International Conference System Modeling and Advancement in Research Trends (SMART), 2019.
- [23] M. R. S. N. A. I. D. D. Md. Tofael Ahmed, "Deployment of Machine Learning and Deep Learning Algorithms in Detecting Cyberbullying in Bangla and Romanized Bangla text: A Comparative Study," in International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAEC), Bhillai, India, 2021.
- [24] Bashar, M. A., Ahmed, M. T., Syduzzaman, M., Ray, P. J., & Islam, A. T. (2014). Text-independent speaker identification system using average pitch and formant analysis. International Journal on Information Theory (IJIT), 3(3), 23-30.
- [25] Ahmed, Md.Tofael, Akter, Nahida, Rahman, Maqsuder, Das, Dipankar, A.Z.M Touhidul, Rashed, Golam. " MULTIMODAL CYBERBULLYING MEME DETECTION FROM SOCIAL MEDIA USING DEEP LEARNING APPROACH," in International Journal of Computer Science and Information Technology (IJCSIT), vol.15, p. 27-37, 2023.

# SE-RESNET: Monkeypox Detection Model

Krishnan Thiruppathi<sup>1</sup>, Selvakumar K<sup>2</sup>, Vairachilai Shenbagavel<sup>3</sup>

Information Technology-Faculty of Engineering and Technology,  
Annamalai University Chidambaram, Tamilnadu 608001<sup>1,2</sup>

Computer Science and Engineering, VIT Bhopal University, Bhopal, Madhya Pradesh, India<sup>3</sup>

**Abstract**—The monkeypox virus, a species of the Orthopoxvirus genus within the family Poxviridae, is answerable for inflicting monkeypox. The symptoms of monkeypox last for about two to three weeks, which is often a self-limiting infection. There may be extreme cases. Recently, the case fatality rate has been in the region of 3-6. When developing a clinical medical diagnosis, it is vital to incorporate different rash diseases such as pox, measles, bacterial skin infections, scabies, syphilis, and medically connected allergies. Pathology at the symptom stage of the sickness could aid in distinctive monkeypox from chickenpox or smallpox. The dataset's machine learning model should not be used for clinical diagnosis, but rather for developing a new model to identify illness fast. The gray scale versions of the original photos in the Monkeypox grey file could make it easier to figure out training more quickly. The channel-wise feature responses that are adaptively re-calibrated are handled by the "Squeeze-and-Excitation" (SE) block. To do this, cross-channel dependency must be explicitly modeled. To demonstrate how these architectures are put together and how these building pieces may be layered to produce SE-Resnet designs in monkeypox image sets that generalize very well. Also, demonstrate that employing SE blocks significantly enhances the performance of current state-of-the-art CNNs while incurring just a little computational cost.

**Keywords**—Squeeze-and-Excitation (SE); monkeypox; poxviridae; prodromal; chickenpox; prodromal

## I. INTRODUCTION

As a result of the outbreak of COVID-19 in 2020, the whole globe was put in jeopardy; however, the emergence of monkeypox in 2022, which was reported by a number of countries, reveals the existence of yet another global danger. The Zoonotic Orthopoxvirus is responsible for the infectious illness known as monkeypox. The virus that causes monkeypox may be a member of the Poxviridae family (a member of the genus Orthopoxvirus) and is closely associated with each chickenpox and smallpox. However, transfer from person to person is also highly prevalent [1]. Rats and monkeys are the major disease transmission vectors. The virus was first found in an exceedingly monkey's body by researchers in a facility in Copenhagen, Denmark, in 1958 [17]. In 1970, amid a significantly additional aggressive plan to eliminate smallpox, the Democratic Republic of the Congo reported the primary human case of monkeypox [19]. This happened all during the campaign. Many people who live in close proximity to tropical rainforests are susceptible to contracting monkeypox, which is often spread across the central and western regions of Africa. By having direct touch with an infected animal, person, or object, a person can catch the virus. Direct body-to-body contact, animal bites, respiration

droplets, or mucus from the eyes, nose, or mouth are all approaches that it would spread [18].

Fever, body pains, and exhaustion are some of the early-stage symptoms that people who have been infected with monkeypox may experience. The long-term impact of monkeypox is a red bump that appears on the skin [5]. In 1996, several villages in Zaire's Kasai Oriental region reported receiving instances of monkeypox, according to the Katako-Kombe Health Zone. These communities existed inside its boundaries. For instance, the Democratic Republic of the Congo. In conjunction with the Centers for Disease Control and Prevention (CDC), the World Health Organization (WHO) appeared into this incidence. They extracted MPV from the lesions of active patients after identifying 92 likely instances that first appeared between February 1996 and February 1997. Between February 1996 and February 1997, all of the cases got started. In response to the continued reporting of instances, the World Health Organization and the Centers for Disease Control and Prevention (CDC) started a fresh inquiry in October 1997. The field investigation's findings, which are summarized in this article, show that the current monkeypox outbreak is the most serious one ever recorded in humans [2]. Phylogenetic analysis imply that the virus has been circulating outside of locations where it has been prevalent for some time without being recognized, perhaps disguising itself as other sexually transmitted illnesses (STIs) [20]. Individuals with a polymerase chain reaction-verified illness who were identified in sixteen different countries across five continents during the months of late April and late June 2022 were evaluated for probable exposures, demographic features, clinical findings, and outcomes. Although it was not possible to prove it, sexual transmission was thought to have occurred in 95% of patients but could not be proven. The data from 23 people were used to find that seven days was the median length of time spent in the incubation phase. In all, 13% of the people who contracted the illness were admitted to the hospital, most often for pain treatment. There were no recorded fatalities [4]. There is evidence that monkeypox can be passed from humans to their dogs based on the timing of the start of symptoms in both the human patient and their dog after the human patient became infected. The dog had sores on its skin and mucosa, and a test for monkeypox virus was positive.

PCR records from anal and mouth swabs exhibit that the sickness is existing in puppies and that it is now not simply transferred there via close contact with humans or by means of airborne transmission (or both). The results of this study should spark discussion regarding whether it is necessary to keep people with the monkeypox virus segregated from their

pets [22]. There is a possibility that the monkeypox virus belongs to not one but two distinct families. The West African clade affords a greater wonderful outlook with a case fatality price of much less than 1%. The most hazardous of the groups on the different aspect is the Central Basin clade, additionally referred to as the Central African clade. A case fatality rate of up to 11% may also observe to kids who have no longer acquired their vaccines. A full recovery for the remaining individuals often occurs four weeks following the commencement of their symptoms, with the possible exception of scarring and skin discoloration [21].

Patients often complete their recuperation within this time range. There is presently no treatment that is known to be effective against an infection caused by monkeypox. Treatment for viral infections consists on relieving the patient's symptoms as well as possible. Nevertheless, there are preventative measures that may be performed in order to avert an epidemic. The contaminated person must continue to be isolated, put on a surgical mask, and maintain lesions blanketed as lots as is virtually viable till all crusts on lesions have naturally fallen off and a sparkling pores and skin layer has grown. During this time, the infected person should also keep lesions covered. Research may additionally be achieved to decide the viability of the usage of components that have already been demonstrated to be really helpful in opposition to Orthopoxvirus in animal trials and extreme vaccinia vaccine sequelae in extra intense conditions. It is not known if the intravenous vaccinia immune globulin, the intracellular viral release inhibitor tecovirimat, or the oral DNA polymerase inhibitor brincidofovir will be effective against the monkeypox virus [14].

## II. RELATED WORK

Monkeypox provides challenges to public health officers and healthcare authorities in the areas of surveillance, laboratory capability, disease management, and treatment. Monkeypox cannot be recognised, diagnosed, treated, or prevented from spreading in many countries because to a lack of knowledge and experience in these fields. Disease monitoring systems demand initial and long-term financial and human resources. Mandatory illness reporting has enhanced Disease Surveillance and Response system reporting. Alerts are routine, but diagnostic samples and preventative techniques like contact tracing and patient seclusion are not. Human and animal health sectors must coordinate efforts and share information because monkeypox is a zoonotic disease [3]. Image analysis will be used as a method for training and developing machine learning models to classify the monkeypox illness. In addition to this, a modified VGG16 model is constructed, and it is tested for its capacity to identify between individuals who have monkeypox illness and those who do not. The fact that it was such a huge network in terms of the amount of parameters that needed to be trained was the most significant drawback [13]. In 2003, the Midwest saw an outbreak of monkeypox. Rats that were infected with the disease were acquired from a business that sold exotic pets and residential homes. After being killed, the rats had been taken to the United States Army Medical Research Institute of Infectious Diseases for investigation. Real-time polymerase chain reaction (PCR), enzyme-linked

immunosorbent assays (ELISA), and viral way of life had been used to analyse and prepare rodent tissue samples. For the purpose of identifying monkeypox viral DNA, we created and examined two distinct real-time PCR procedures. These methods used the F3L and N3R genes of the Vaccinia virus as their respective targets. The DNA of the orthopox virus and other bacteria were used to verify the assays. The presence of orthopoxvirus in rodents was shown by electrochemiluminescence (ECL) using panorthopox. Both the specific PCR test and the pan-orthopox test revealed that seven out of 12 (58%) of the animals had monkeypox (in at least one tissue). The outcomes of the PCR and the ECL were different. Both the PCR and the ECL tests came out positive in one hamster and three gerbils. Our team also used immune histology, electron microscopy, and culture on a variety of different cell lines to verify the monkeypox virus's existence. The samples' PCR findings revealed that the Zaire-96-I-16 monkeypox virus was present in each occasion (a human isolate from the Congo). Techniques that can be used to identify orthopox viruses include real-time PCR and ECL. Early detection is essential for both naturally occurring outbreaks and bioterrorism due to recent viral transmissions of the monkeypox virus [12]. Following an experimental MPXV infection, squirrels were observed for both the severity of the disease's clinical signs and the amount of virus that was expelled from their bodies through their body fluids. This was carried out to ascertain whether the virus could spread from an animal to a human. The outcomes of this study revealed that while some rope squirrels were unable to recover from their very minor illnesses, others were. They can expel a sizable amount of the virus through their lips, nostrils, eyes, and bowel movements. This information aids epidemiologists and public health experts in their understanding of the potential risks that interacting with rope squirrels may provide to local communities in Africa. Disease ecologists will additionally advantage from this discovery considering that it will assist them recognize how the MPXV virus is maintained and transferred from animals to humans [8].

When employing the MPXV PCR technique to screen for gonorrhoea and chlamydia, samples collected from four different males revealed the presence of monkeypox virus (MPXV) DNA. Through the use of serology, it was discovered that all three people had been subjected to MPXV, and the virus was successfully cultured from samples taken from two of the patients. These findings suggest that some incidences of monkeypox have not yet been identified, and they signal that testing and quarantining those who report symptoms might not be adequate to prevent the breakout of the illness [7]. However, it might also be investigated a wide variety of antiviral pills that had been first created for the remedy of smallpox and different viral infections [6]. There are no specific medications approved to treat monkeypox virus infection at this time. Cases in the continuing monkeypox pandemic in 2022 have been discovered to have peculiar clinical features. They consist of the absence of prodromal symptoms (such as lymphadenopathy and fever) and a propensity for early lesions to develop on the vaginal and perianal areas of the body. This methodology's novelty lies in its comprehensive clinical diagnosis approach, considering various similar diseases when identifying monkeypox. It

introduces a rapid, precise machine learning model for illness detection, especially monkeypox. It enhances efficiency by using grayscale images and “Squeeze-and-Excitation” (SE) blocks for Convolutional Neural Networks (CNNs) in medical image analysis, all while maintaining diagnostic accuracy. Combining clinical insights, machine learning, and innovative image processing, this approach has the potential to advance healthcare diagnostics and patient care. The treatment of symptoms will vary according to the systems involved or the individual syndromes. Patients who are at a high risk of developing severe illness or who already have high-risk disease characteristics may benefit from antiviral medication that alleviates severe disease and lowers risk. We are the guardians of the body that we inhabit. It is imperative that we take the necessary steps to adopt healthy living habits in order to forestall, mitigate, or otherwise take control of diseases and illnesses [16]. We used the SQUEEZE- AND- EXCITATION-Resnet layer of a convolutional neural network to find the monkeypox disease. This reduces the high risk of getting a serious illness.

### III. METHODS

The performance of any basic design may be enhanced by using this simple yet efficient add-on module, and the extra computational weight is only slightly increased by doing so. Squeeze-and-Excitation (SE) blocks were used in this approach to calculate the channel attention. The effect of squeeze- and- excitation (SE) blocks on traditional architectures will subsequent be examined, alongside with SE blocks’ overall performance in a range of computer vision applications. In present day designs of convolutional neural networks, the frames are equal to the channels in a tensor produced by using a convolutional layer. Typically, the dimensions of this tensor are (B, C, H, W), the place B stands for the batch size, C for the channels, and H and W for the corresponding spatial dimensions of the feature maps. In different words, the notation (B, C, H, W) may additionally be used to point out the dimensions of this tensor (H represents the height and W represents the width). Convolutional filters were used to extract a range of properties from the input data, which led to the creation of channels. In spite of this, there is a distinct possibility that the channels may not all possess the same degree of representational value. Because it’s likely that certain channels are more important than others, it’s a good idea to give each one a weight that’s proportionate to its value before the information is passed on to the next layer. This will ensure that the most important information gets sent.

#### A. Channel Attention

In a convolutional neural network, the two primary components are as follows:

- The dimensions provide a representation of the input tensor, which is typically a four-dimensional tensor (B, C, H, W).
- The weights for each layer are stored inside the trainable convolutional filters.

The convolutional filters are the ones that are in charge of generating the feature maps, and they do this by basing those maps on the learned weights that are contained inside

those filters or to put it another way, the feature maps are constructed by the convolutional filters. Together, these filters learn several feature representations of the target class data that the input tensor includes in the image. While other filters are taught to learn textures, some are taught to learn edges. Therefore, the variety of channels determines the range of convolutional filters that are used to study the distinctive feature maps of the input. These feature maps also have varied degrees of usefulness, based on what we know about frame selection in photography at this point in time. This shows that certain feature maps have high value than others do. A feature map that learns background texture transitions, for example, can be less useful and required for the learning process than a feature map that learns edge information since the latter already has the edge information. As a direct result, it is essentially suitable to supply the extra significant feature maps with a higher degree of relevance in contrast to the related feature maps. This lays the groundwork for how attention will be directed. We prefer to focal point our “attention” on the channels that supply the best significance, which, in practise, implies that we favor to prioritise some channels above others and provide them extra importance. The most straightforward strategy for achieving this objective is to apply a bigger scaling factor to the channels that carry a greater amount of significance. The Squeeze-Excitation Networks theory makes the exact same prediction about the likelihood of this happening.

#### B. Squeeze-and-Excitation Blocks

An architectural element called the Squeeze-and-Excitation Block permits dynamic channel-wise feature recalibration, which boosts a network’s representational power. This was done by the network being able to squeeze and stimulate information [11]. The procedure is as follows:

- As an input, the block makes use of a convolutional block.
- When employing average pooling, each channel is “squished” into a single numerical number.
- A dense layer that is then followed by a ReLU will increase non-linearity, and the complexity of the output channel will be lowered by a ratio.
- A sigmoid follows another thick layer, creating a smooth gating function for each channel.
- Last but not least, we assign a weight based on the side network, often referred to as the “excitation,” to each feature map of the convolutional block. Fig. 1 and 2 explains the Squeeze-and-Excitation Block.

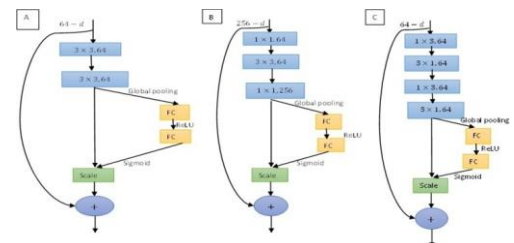


Fig. 1. (A) Rudimentary SE-ResNet core module (B) Congestion SE-ResNet module. (C) Trifling SE-ResNet module.



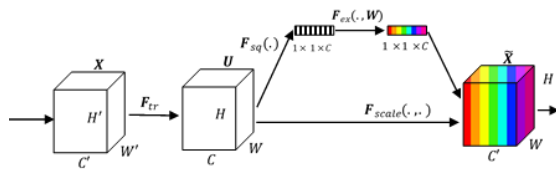


Fig. 2. Squeeze-excitation module.

- The SE-block, also known as the Squeeze-and-Excite block, is a straightforward plug-in module made up of three components:
- Squeeze Module
- Excitation Module
- Scale Module

1) *Squeeze module*: To maximise channel attention, it would be preferable if the feature maps' scale should be modified to fit the maps themselves. As a consequence, the best channel attention would be obtained. In a nutshell, the output tensor from a convolutional layer is the feature map set. The letters B, C, H, and W, in the well-known tensor notation stand for the batch size, channels, height, and width of the feature mappings (B, C, H, W). For the purpose of simplicity, let's think of it as a three-dimensional tensor of type (C, H, W). In essence, what things is the depth (the quantity of channels or feature maps contained in the tensor) and geographic dimensions of every feature map. We must be concerned with HW pixels (or values) as a whole in order to make the attention paid to channels flexible to each channel taken independently. In order to make the interest genuinely adaptable, this would effectively suggest that you would be dealing with a whole of C, H, and W variables. This is due to the previous statement. Given that the number of channels in modern neural networks increases proportionally to the network depth, this figure will grow to be quite large. Therefore, in order to simplify the computation requirements of the whole operation, the usage of a feature descriptor that may condense the data included in each feature map to a single value is required [10]. The Squeeze Module was created as a result. Convolutional neural networks often make use of the pooling approach to minimize the quantity of space that the features occupy, even if the spatial dimensions of the feature maps can also be decreased to a single cost utilizing a range of distinct feature descriptors. Both the maximum pooling method and the average pooling method are widely used methods of pooling. While the second approach gets the greatest pixel value inside the same specified frame, the first method determines the average pixel values within a given window. Both have advantages and disadvantages in proportion to how beneficial they are overall. Although max pooling performs a decent job of protecting the pixels that are most likely to activate, it also has the potential to be highly noisy and ignores the neighboring pixels. Despite no longer maintaining the information, common pooling creates a smoother average of all the pixels covered inside that window

[15]. As shown in Fig. 3, we conducted an ablation inquiry to assess each descriptor's performance: Global Average Pool (GAP) and Global Max Pool (GMP). By averaging out each and every pixel that makes up the feature map, the Global Average Pool (GAP) process, which is employed via the Squeeze Module, essentially compresses the entire feature map to a single value. The Global Average Pool (GAP) operation allows for this. This option is chosen because, in contrast to the other two options, it produces a less chaotic atmosphere. As a consequence, if the input tensor is (CHW), the GAP operation will be performed, yielding an output tensor of shape (Cx1x1), which is simply a vector of length C with every feature map decreased to a single value. The output tensor that is created after it has been subjected to the GAP operator has the shape (Cx1x1) after doing so. In addition, the application of the GAP operator to the input tensor will determine the shape of the output tensor (Cx1x1).

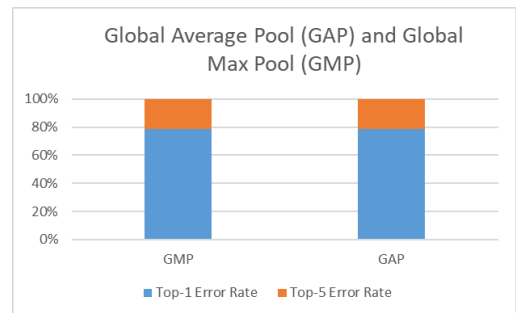


Fig. 3. Global Average Pool (GAP) and Global Max Pool (GMP).

The findings of our comparison between the Squeeze version and the No-Squeeze variant are shown in the table that follows. They did this so they could evaluate the significance of the Squeeze operator. Fig. 4's No-Squeeze variant, which demonstrates that the tensor protecting the feature maps was once no longer condensed to a single pixel and that the Excitation module worked on the full tensor instead of simply a component of it, serves as an illustration of this idea [9].

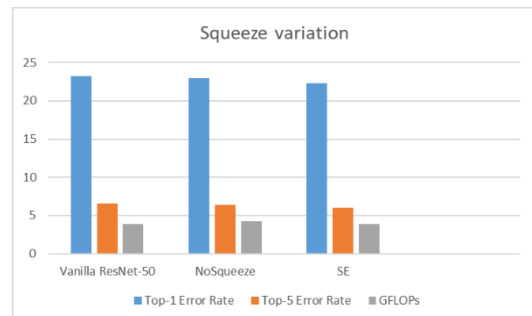


Fig. 4. Squeeze variation.

2) *Excitation module*: The second stage of the module includes gaining knowledge of the adaptive scaling weights for each of these channels after the input tensor used to be reduced in size to a good deal greater sensible measurement of (Cx1x1). We locate that the first-rate approach for mapping the

scaling weights for the Excitation Module, which is existing in the Squeeze-and-Excitation Block, is a fully connected Multi-Layer Perceptron (MLP) bottleneck structure is shown in Fig. 5. The input and output layers and one hidden layer combine to form this MLP bottleneck. The three levels all have the same form. As a reduction block, the hidden layer takes the input space and reduces it in line with the reduction factor to a more manageable size (which is set at 16 by default). Then, in order to use it as the input tensor, the previously compressed region is inflated back to its original size. The following three factors may additionally be used to summarily illustrate the adjustments in dimensionality that take region at every layer of the MLP:

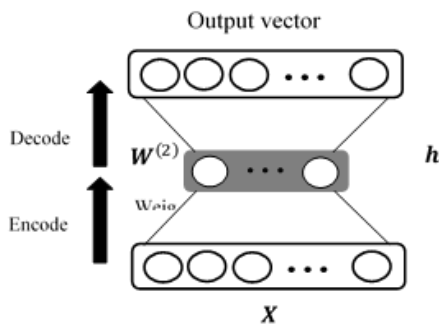


Fig. 5. Multi-Layer Perceptron (MLP) structure.

- The form that the input takes is  $(C \times 1 \times 1)$ . As a direct consequence of this, the input layer has some neurons of type  $C$ .
- This is reduced by a factor of reduction denoted by the letter  $r$  in the buried layer, which brings the total number of neurons to  $C$  raised to the power of  $r$ . When the output is projected again into the equal dimensional area as the input, the whole number of neurons grows to  $C$ .
- The last stage involves projecting the output back into the same dimensions space as the input.

In conclusion, the output is a weighted version of the same tensor with the same form, and the input is a tensor with the form  $(C \times 1 \times 1)$  that is used as the input. Fig. 6 displays the outcomes of their studies on how a SE module integrated into ResNet-50 architecture functions while utilising a variety of reduction ratios.

In a perfect world, the value of  $r$  would be set to 1, which would transform the network into a square that is totally linked on all levels and maintains the same width throughout. This would lead to improved information transfer as well as more interaction across channels (CCI). However, there is a trade-off that can be made between increasing the complexity of the system and enhancing its performance with a lower  $r$ . The trade-off may be made in any direction. As a consequence of this, we reach the conclusion that the default figure for the reduction ratio should be 16, and we base this conclusion on the Fig. 6 was shown before. This is a hyper parameter that can be adjusted further in order to achieve a higher level of performance. There is room for

more adjustment.

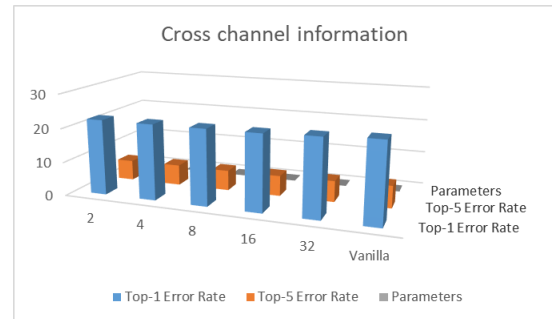


Fig. 6. Cross channel information

3) *Scale module*: Initial processing of the “excited”  $(C \times 1 \times 1)$  tensor is carried out with the aid of a sigmoid activation layer, which limits the values to a range that lies between zero and 1. This is done after the tensor has been acquired via the Excitation Module. Following that, the output is right away utilized to the input the usage of a basic broadcasted element-wise multiplication. Each channel or feature map in the input tensor is scaled the usage of the splendid learned weight from the MLP in the excitation module. This is done in the subsequent phase. More study was once carried out on ablation, with a unique center of attention on the results of a number of non-linear activation functions that might also be used as the excitation operator. In Fig. 7, the research’s conclusions are displayed. Since we conclude from the data that it gives the highest level of performance, the sigmoid activation function is chosen as the scale module’s default excitation operator. In conclusion, the Global Average Pooling (GAP) algorithm is used through the Squeeze Excitation Block (SE Block) to limit an input tensor with the shape  $(C \times H \times W)$  to a tensor with the shape  $(C \times 1 \times 1)$  earlier than the Multi-Layer Perceptron (MLP) bottleneck structure receives the  $C$ -length vector and makes use of it to create a weighted tensor with the identical shape  $(C \times 1 \times 1)$ . The inner spatial convolution of the block is accompanied by way of the squeeze-excitation block, which takes place earlier than the last convolutional layer. As a result, rather than functioning as the add-on that was initially planned, it now functions more like an integrated component. The default configuration, which covered including a SE-block after the eleven convolutions that have been performed, was once eventually chosen by way of SE-Net in spite of the reality that they had carried out ablation experiments and evaluated this integration technique. The most recent method for distinguishing monkeypox from other illnesses by analysing the images in the dataset is state-of-the-art (SOTA) detection utilising effective nets. The importance of being aware of the channels being used as well as the strength of Squeeze Excitation blocks are highlighted by this.

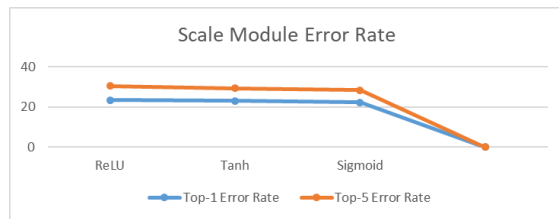


Fig. 7. Scale module.

#### IV. PERFORMANCE EVALUATION

The full set of test findings is investigated and analysed the use of the statistical methods now used by means of the majority of researchers, inclusive of accuracy, precision, recall, F1-score, sensitivity, and specificity. Due to the small range of lookup participants, the typical statistical effects are mentioned as a self-assurance interval with a 95% level of significance. This is followed by previously published research that also used a small dataset. Monkeypox may be classified for the purposes of our dataset as either true positive (Tp) or true negative (Tn), depending on how accurately people are diagnosed; alternatively, it may be classified as either false positive (Fp) or false negative (Fn), depending on how accurately people are diagnosed. Fig. 8 displays the True positive rate and False positive rate AUC scores. Fig. 9 displays the testing accuracy, validation accuracy, and testing accuracy with respect to the number of epochs. In this comparative analysis of algorithms for the detection of monkeypox disease, the focus was on evaluating their performance in conjunction with the SE-Resnet architecture. The dataset used for this evaluation consisted of a total of 228 images, with 102 of them falling into the 'Monkeypox' category, while the remaining 126 represented cases of 'Others,' encompassing diseases like chickenpox and measles.

Several methodologies and algorithms were assessed in terms of their accuracy in distinguishing monkeypox cases from others. Notable among these was the application of pre-trained deep learning (DL) models, as outlined in the study by Sitaula and colleagues in 2022. In this approach, these pre-trained models were fine-tuned using custom layers, and their performance was meticulously analyzed using four well-established metrics. This method yielded an accuracy rate of 87.13%, indicating its effectiveness in monkeypox detection.

Other algorithms included widely recognized deep convolutional neural networks (CNNs), such as ResNet-18 and GoogLeNet, which had 73.33% and 77.78% accuracy, respectively. Furthermore, more complex models like EfficientNet-B0 with 91.11% accuracy and NasnetMobile with 86.67% accuracy demonstrated their capability in monkeypox detection. Shuffle Net, MobileNetv2, CNN (with 3 layers), and LSTM (with 3 layers) also underwent evaluation, with accuracy percentages of 80%, 91%, 64%, and 94%, respectively. These results displayed the varying degrees of success in identifying monkeypox using different architectures.

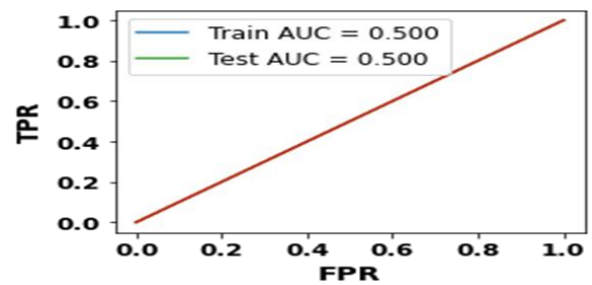


Fig. 8. True positive rate vs False positive rate.

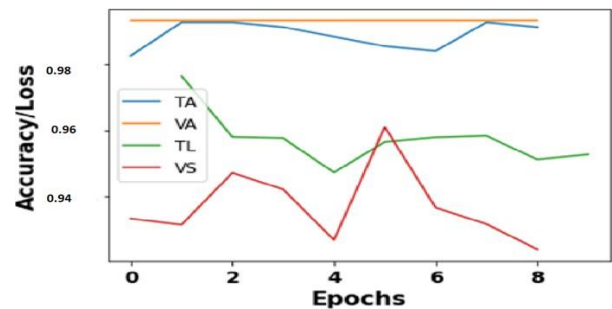


Fig. 9. Accuracy/Loss vs number of Epochs.

However, the most promising performer in this analysis was the proposed SE-Resnet architecture. SE-Resnet leverages the Squeeze-and-Excite block, composed of three critical components: the Excitation Module, Squeeze Module, and Scale Module. This innovative approach demonstrated an exceptional accuracy rate of 96%. The SE-Resnet architecture outperformed all other methods, highlighting its efficacy in enhancing the accuracy and reliability of monkeypox disease detection. In summary, this comprehensive comparison underscores the significance of the SE-Resnet architecture in achieving a remarkable accuracy rate of 96% in monkeypox detection, thereby offering a promising avenue for improving the efficiency of disease diagnosis and potentially enhancing patient care in the field of healthcare.

#### V. CONCLUSION

In this approach, we created a unique dataset for machine learning model training and development that may be utilised to categorise the monkeypox illness using image analysis techniques. The inadequacy of older architectures to accurately characterise channel-wise feature dependencies is something that is made clearer by the introduction of SE blocks. We have high hopes that this new information will be helpful in the categorization of monkeypox images, which calls for highly discriminative characteristics. In addition, a version of the SE-ResNET model is built, and its capacity to distinguish between patients who have and do not have monkeypox illness is investigated in two distinct investigations. Our suggested model, SE-ResNET, was able to attain an accuracy of around 95% with a score of 0.5% AUC. Some of the boundaries of our work can be overcome by means of consistently accumulating new images of monkeypox-infected patients, updating the dataset, checking out the overall performance of the proposed SE-ResNET model on highly skewed data, evaluating the overall performance of our model,

and the use of the proposed model to construct mobile-based prognosis tools.

#### ACKNOWLEDGMENT

I would like to express our sincere gratitude to all individuals and organizations who have contributed to the successful completion of this research article. First, I extend my deepest appreciation to our research advisor Dr. K. Selvakumar, Head and Professor, whose guidance, expertise, and continuous support were invaluable throughout the entire research process. Their valuable insights and constructive feedback significantly enhanced the quality and rigor of this study. Furthermore, we would like to acknowledge the contribution of Dr. S. Vairachilai, Dr. R Suban, and Dr. Pasupathy, who provided assistance with data collection, analysis, and interpretation. Their expertise and dedication played a crucial role in the successful execution of this research. Once again, I express my sincere gratitude to everyone involved, and we hope that our research contributes to the advancement of knowledge in this field. Feel free to modify and adapt this acknowledgment text to suit your specific circumstances and the individuals and organizations you wish to acknowledge in your journal article.

#### REFERENCES

- [1] Alakunle, E., Moens, U., Nchinda, G., and Okeke, M. I 'Monkeypox virus in Nigeria: infection biology, epidemiology, and evolution' *Viruses*, 2020, Vol.12, No.11, pp.1257.
- [2] Aplogan, A., and Szczeniowski, M. 'Human monkeypox-Kasai Oriental', *Democratic Republic of... MMWR: Morbidity & Mortality Weekly Report*, 1997, Vol.46, No.49, pp.1168-1171.
- [3] Bass, J., Tack, D. M., McCollum, A. M., Kabamba, J., Pakuta, E., et al. 'Enhancing health care worker ability to detect and care for patients with monkeypox in the Democratic Republic of the Congo', *International health* 2013, Vol. 5, No.4, pp.237-243.
- [4] Benites-Zapata, V. A., Ulloque-Badaracco, J. R., Alarcon-Braga, E. A., Hernandez-Bustamante, E. A., et al. 'Clinical features, hospitalisation and deaths associated with monkeypox: a systematic review and meta-analysis', *Annals of clinical microbiology and antimicrobials* 2022, Vol. 21, No.1, pp.1-18.
- [5] CDC: Clinician Outreach and Communication Activity (COCA): What Clinicians Need to Know about Monkeypox in the United States and Other Countries. CDC website. Reviewed May 20, 2022
- [6] CDC: Monkeypox: Healthcare Professionals: Isolation and Infection Control at Home. CDC website. Up-dated August 11, 2022. Accessed August 23, 2022. <https://www.cdc.gov/poxvirus/monkeypox/clinicians/infection-control-home.html><https://www.cdc.gov/poxvirus/monkeypox/clinicians/infection-control-home.html>
- [7] De Baetselier, I., Van Dijck, C., Kenyon, C., Coppens, J., Michiels, J., et al. 'Retrospective detection of asymptomatic monkeypox virus infections among male sexual health clinic attendees in Belgium', *Nature Medicine*, 2022, pp.1-5.
- [8] Falendysz, E. A., Lopera, J. G., Doty, J. B., Nakazawa, Y., Crill, C., et al. 'Characterization of Monkeypox virus infection in African rope squirrels (*Funisciurus sp.*)', *PLoS neglected tropical diseases*, 2017, Vol.11, No.8, pp. e0005809.
- [9] He, J., and Jiang, D. 'Fully automatic model based on SE-resnet for bone age assessment' *IEEE Access*, 2021, Vol.9, pp.62460-62466.
- [10] Hu, J., Shen, L., and Sun, G. 'Squeeze-and-excitation networks' In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132-7141.
- [11] Jiang, Y., Chen, L., Zhang, H., and Xiao, X. 'Breast cancer histopathological image classification using convolutional neural networks with small SE-ResNet module', *PloS one*, 2019, Vol.14, No.3, pp.e0214587.
- [12] Kulesh, D. A., Loveless, B. M., Norwood, D., Garrison, J., Whitehouse, C. A., et al. Monkeypox virus detection in rodents using real-time 3-minor groove binder TaqMan assays on the Roche LightCycler, *Laboratory investigation*, 2004, Vol.84, No.9, pp.1200-1208.
- [13] Manjurul Ahsan, M., Ramiz Uddin, M., Farjana, M., Nazmus Sakib, A., Al Momin, K., et al. 'Image Data collection and implementation of deep learning-based model in detecting Monkeypox disease using modified VGG16', *arXiv e-prints*, 2022, arXiv-2206.
- [14] McCollum, A. M., and Damon, I. K., 'Human monkeypox', *Clinical infectious diseases*, 2014, Vol. 58, No.2, pp.260-267.
- [15] Mekruksavanich, S., Jitpattanakul, A., Sithithakerngkiet, K., Youplao, P., and Yupapin, P. 'ResNet-SE: Channel Attention-Based Deep Residual Network for Complex Activity Recognition Using Wrist-Worn Wearable Sensors' *IEEE Access*, 2022.
- [16] Monkeypox signs and symptoms. (accessed on may 30, 2022). <https://www.cdc.gov/poxvirus/monkeypox/symptoms.html>, 2022.
- [17] Moore, M., and Zahra, F. Monkeypox. In *StatPearls* [Internet]. StatPearls Publishing, 2021.
- [18] Nguyen, P. Y., Ajisegiri, W. S., Costantino, V., Chughtai, A. A., and MacIntyre, C. R. 'Reemergence of human monkeypox and declining population immunity in the context of urbanization, Nigeria', 2017-2020. *Emerging Infectious Diseases*, 2021, Vol.27, No.4, pp.1007.
- [19] Nolen, L. D., Osadebe, L., Katomba, J., Likofata, J., Mukadi, D., et al. 'Extended human-to-human transmission during a monkeypox outbreak in the Democratic Republic of the Congo', *Emerging infectious diseases*, 2016, Vol.22, No.6, pp.1014.
- [20] O'Toole, A., and Rambaut, A. 'Initial observations about putative APOBEC3 deaminase editing driving short-term evolution of MPXV since 2017', *ARTIC Network*, 2022.
- [21] Sklenovska, N., and Van Ranst, M. 'Emergence of monkeypox as the most important orthopoxvirus infection in humans', *Frontiers in public health*, 2018, Vol. 6, pp. 241.
- [22] Tar'in-Vicente, E. J., Alemany, A., Agud-Dios, M., Ubals, M., Sun'er, et al. 'Clinical presentation and virological assessment of confirmed human monkeypox virus cases in Spain: a prospective observational cohort study', *The Lancet*, 2022, Vol.400, No.10353, pp. 661-669

# Enhancing Skin Cancer Detection Through an AI-Powered Framework by Integrating African Vulture Optimization with GAN-based Bi-LSTM Architecture

N.V. Rajasekhar Reddy<sup>1</sup>, Dr Araddhana Arvind Deshmukh<sup>2</sup>, Dr. Vuda Sreenivasa Rao<sup>3</sup>, Dr Sanjiv Rao Godla<sup>4</sup>,  
Prof. Ts. Dr. Yousef A.Baker El-Ebiary<sup>5</sup>, Liz Maribel Robladillo Bravo<sup>6</sup>, R. Manikandan<sup>7</sup>

Professor, Department of Information Technology, MLR Institute of Technology, Hyderabad<sup>1</sup>

Head and Associate Professor, Department of Artificial Intelligence and Data Science, Marathwada Mitra Mandal College of Engineering-Affiliated to Savitribai Phule Pune University<sup>2</sup>

Associate professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India<sup>3</sup>

Professor, Department of CSE (Artificial Intelligence & Machine Learning), Aditya College of Engineering and Technology-Surapalem, Andhra Pradesh, India<sup>4</sup>

Faculty of Informatics and Computing, UniSZA University, Malaysia<sup>5</sup>

Universidad César Vallejo, Peru<sup>6</sup>

Research Scholar, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai-600062, Tamil Nadu, India<sup>7</sup>

**Abstract**—One of the more prevalent and severe cancer kinds is thought to be skin cancer. The main objective is to detect the melanoma in initial stage and save millions of lives. One of the most difficult aspects of developing an effective automatic classification system is due to lack of large datasets. The data imbalance and overfitting problem degrades the accuracy. In this proposed work, this problem can be solved using a Generative Adversarial Network (GAN) by generating more training images. Traditional RNNs are concerned with overcoming memory constraints. By using a cyclic link on the hidden layer, these models attain Long short-term memory. However, RNNs suffer from the issue of the gradient disappearing, which affects learning performance. To overcome these challenges this work proposes Bidirectional Long Short-Term Memory (Bi-LSTM) deep learning framework for skin cancer detection. The dataset which is collected from the International Skin Imaging Collaboration were used in image processing. A novel metaheuristic enthused by the routine of African vultures is proposed in this proposed work. The African Vulture Optimisation Algorithm (AVOA) algorithm is designed to select optimum feature of skin image. The accuracy of the proposed method obtains 98.5%. This comprehensive framework, encompassing GAN-generated data, Bi-LSTM architecture, and AVOA-based feature optimization, contributes significantly to enhancing early melanoma detection.

**Keywords**—Skin cancer; generative adversarial network; Bi-LSTM; African Vulture Optimisation (AVO); deep learning (DL)

## I. INTRODUCTION

Skin cancer has become difficult to diagnose because of apparent similarities. Melanoma is the well-known type of skin cancer, have been caused for a significant number of deaths in recent years. Recent surveys show that in contrast

with various cancer types, the number of skin cancer patients is growing annually. Melanocytes, the skin external cells, are affected by it. It has several cell types that result in the skin becoming darker. It irregularly comes in a variety of dark tones. It can also be noticed on the skin in colorless or in shades of rosy pink in color, royal purple, azure, and more. It is more deadly and hazardous because it spreads quickly. Melanoma can be discovered everywhere on the human body, despite the fact that it typically develops on the lower limb's backside [1]. According to data from the World Health Organization (WHO), there are several thousand cases worldwide with a high risk of death by the year 2020. According to that, 324,635 new cases have been reported globally, of which 57,043 have resulted in death. The researcher also demonstrate that, of every 100 persons with melanoma, 18 will not survive [2]. Melanocytes in the epidermal layer have the potential to produce excessive amounts of melanin at a high rate in several circumstances. For instance, melanin is produced when strong UV radiation from sunlight is exposed for an extended period of time. Melanoma, a deadly kind of skin cancer, is the outcome of melanocytes' unusual growth. For successful treatment of melanoma, an early diagnosis is crucial. The survival percentage for 5 years is approximately 92% if the skin cancer is sensed at an earlier phase [3]. Accurate and timely diagnosis of melanoma, a deadly form of skin cancer, remains a pressing challenge in the field of dermatology and healthcare. Despite advancements in medical technology, the current diagnostic methods are often subjective and error-prone, leading to delayed diagnoses and poorer patient outcomes. The need for a more reliable and efficient approach to melanoma detection is evident, especially considering the increasing incidence of skin cancer worldwide. This research endeavors to address

these critical issues by developing an automated melanoma detection system that leverages the power of deep learning and convolutional neural networks.

Historically, skin cancer diseases have been diagnosed and detected by manual examination and inspection by sight. These methods for skin doctor to visually assess and screen lesion photographs are time-consuming, difficult, and error-prone [4]. Use of the ABCDE rule, which stands for asymmetry, boundaries, color, diameter, and evolving, is a typical way for spotting melanomas [5]. To diagnose melanoma, these warning indicators are monitored. The first warning indicator is a mole that is very asymmetrical or has uneven border patterns, as well as one that is larger than 6 mm in diameter and has an odd color. All of these indicators are tracked in order to study how researcher changes over time which determines the presence of melanoma. This methodology could be inaccurate and prone to measurement mistakes [6]. Effective feature extraction, classifiers, and color capture are required for the detection of skin lesion images. Recent developments have made it possible to diagnose melanoma accurately by molecular dermatopathology, which necessitates a discussion between a pathologist and a dermatologist. Even though the majority of dermatopathologists can determine the histological evaluation of melanocytic lesions by performing a traditional microscopic examination, some melanocytic neoplasms known as typical melanocytic proliferations require expert opinion before categorized as benign or malignant. The investigation of the histopathological features in relation to the clinical and microscopy data is also required because of this. If molecular diagnostics is used incorrectly to decide whether a condition is benign or malignant, it may also be deceptive and lose some of its greatest value [7]. However, this method can only be executed effectively by skilled medical specialists. These complications boost the scientific people to generate innovative systems for melanoma visualization and diagnosis. Melanoma cancer is diagnosed with the use of a computer-aided diagnosis (CAD) system. CAD diagnostic tool evidence can be utilized as a backup diagnosis for melanoma malignancy [8]. The expanding use of machine learning and AI in the disciplines of medical and health care has attracted a lot of study attention in recent years [9]. Clinical image of skin lesions with the existence of artefacts such as hair, veins, and texture and other issues might be challenging to identify melanoma lesions away from non-melanoma lesions. Consequently, the need for imagery preparation is crucial [10]. Early detection has been proven to significantly improve patient survival rates, with a five-year survival rate of approximately 92% when melanoma is detected at an earlier stage. By addressing the shortcomings of current diagnostic methods, this research aims to contribute to saving lives and reducing the burden of melanoma on individuals and healthcare systems.

Convolutional neural networks with deep learning recently entered the field of image-based skin cancer diagnosis and demonstrated diagnostic performance comparable to that of dermatologists. It would be ideal if doctors had assistance in the diagnosis of difficult-to-diagnose melanomas with unique localizations and uncommon subtypes [11]. The segmentation

is a vital step in creating an automated melanoma detection system. However, when there is no variation in image contrast or when there are just slight variations in illumination in the image content, the region of interest or thresholding-based algorithms perform well [12]. ResNet-50 Convolutional Neural Network (CNN) architecture was designed. CNN model employed a varied learning rate for each CNN layer. To slow down learning rates, novel techniques based on the cosines function is applied. The success percentage for the categorization had a sensitivity rate of 82.3% [13]. In deep CNN SoftMax classifier were used. This approach worked effectively for lesion images that were blurry and had many sizes. This method was calculated using skin lesion samples from PH2 and ISIC 2018 and yielded respectable accuracy and dice coefficients of 95% and 93%, for each [14].

The existing diagnostic methods for melanoma are beset by several limitations that hinder their effectiveness. Manual examination, reliant on visual inspection, is not only time-consuming but also susceptible to human subjectivity. The widely-used ABCDE rule, which assesses asymmetry, boundaries, color, diameter, and evolution, can introduce errors due to its reliance on qualitative observations. The presence of noisy images with artifacts like hair, veins, and variations in color makes accurate diagnosis even more challenging. These limitations underscore the need for a more objective, automated, and robust approach to melanoma detection. And also often struggle to accurately detect and segment melanoma lesions due to factors such as inadequate image segmentation, vanishing/exploding gradients in deep learning architectures, and difficulty in handling noisy images with variations like hairs or color changes. The current landscape of melanoma diagnosis is characterized by a conspicuous gap between the pressing need for accurate and timely detection and the limitations of existing methods. While deep learning and CNNs have shown immense potential in image-based tasks, their application to dermatology, especially for melanoma detection, remains a relatively unexplored territory. This research aims to bridge this gap by introducing an innovative approach that combines advanced technology with the critical domain of skin cancer diagnosis. Moreover, certain techniques suffer from computational intensity during feature extraction, leading to time-consuming analyses. In this work AI driven African Vulture optimization with GAN based Bi-LSTM deep framework for melanoma detection is projected. The primary contributions of the suggested work include:

- **Data Collection and GAN Augmentation:** The utilization of skin cancer images from the International Skin Imaging Collaboration (ISIC 2018) addresses the challenge of limited data availability. The integration of Generative Adversarial Networks (GANs) to generate additional synthetic images serves as a solution for data imbalance, enhancing the diversity and size of the dataset.
- **Advanced Preprocessing Techniques:** The implementation of Contrast Limited Adaptive Histogram Equalization (CLAHE) and Weiner filtering on input images aids in enhancing image quality and reducing unwanted artifacts such as hair, veins, and

noise. This preprocessing pipeline ensures that the subsequent analysis is based on clean and standardized data.

- **Precise Skin Contour Segmentation:** The application of Kapur thresholding for skin contour segmentation generates binary images that precisely isolate the skin lesions from the background. This step is pivotal in isolating the region of interest and improving subsequent feature extraction.
- **Optimal Feature Selection:** The introduction of African Vulture Optimization (AVO) algorithm for selecting optimal features from skin images adds a novel contribution. This technique helps to streamline and enhance the feature extraction process, potentially leading to more effective and efficient classification.
- **Hybrid CNN-Bi-LSTM Architecture:** The design of a Convolutional Neural Network (CNN) with Bidirectional Long Short-Term Memory (Bi-LSTM) layers offers a robust architecture for detecting melanoma. This combination enables the model to capture both spatial features from images and sequential patterns, potentially improving accuracy and diagnostic capabilities.

Overall, the key contributions encompass data augmentation, advanced preprocessing, precise segmentation, innovative feature selection, and a sophisticated deep learning architecture, collectively aiming to boost the accuracy and effectiveness of early melanoma detection and diagnosis. The leftover portion of this work is organised as follows: Section II contains comparable work as well as a thorough examination of them. Section III contains information about the problem statement. The proposed AVO-CNN architectures are discussed in detail in Section IV. In Section V, the outcomes of the experiments are presented and examined, and a full comparison of the suggested strategy to current best practises is given. Section VI, is the final section, where the paper is concluded.

## II. RELATED WORKS

Albert [15] proposed the combination of Predict-Evaluate-Correct K-fold (PECK) algorithms and Synthesis and Convergence of Intermediate Decaying Omnigradients (SCIDOG). MED-NODE data set were used to diagnose melanomas via digital image analysis. In the PECK algorithm 153 non-dermoscopic images of lesion were deep ensembled. On that data the state-of-the-art methods were educated and estimated. Considerable improvement in diagnostic performance over the most effective earlier approaches was achieved through introspective learning of the PECK ensemble to increase precision from sparse but high-dimensional training data.

Jiang, Li, and Jin [16] introduced a light-weight based deep learning outline called DRANet to distinguish between 11 dissimilar categories of skin diseases using a factual histopathology data set amassed over the past ten years. DRANet outperforms baseline models (such InceptionV3, ResNet50, VGG16, and VGG19) by a significant margin with

identical parameter sizes and comparative accuracy with fewer parameters. Vanishing/exploding gradients are a concern, and several stacked layers can occasionally have substantial training error. Shorfuzzaman [17] suggested a comprehensible ensemble stacked structure with CNN for early malignance skin tumor detection. Multiple sub-models of CNN which performs the similar classification are collected in the loading ensemble outline, which employs the transfer learning idea. All of the predictions from the sub-models are combined into a novel model termed a meta-learner, which produces the final prediction outcomes. Stacking ensemble model was trained and validated using skin lesions Kaggle dataset obtained over the International Skin Image Collection. The dataset includes 1497 and 1800 pictures of benign and cancerous moles, respectively. According to evaluation results, the ensemble model has better accuracy of 95.76%, sensitivity of 96.67%, and AUC of 0.957. Low-quality prediction results from inadequate image segmentation.

Wei, Ding, and Hu [18] suggested a simple model for detecting skin cancer that uses feature discrimination and the fine-grained classification principle. Two sets of training samples of the recognition model are first introduced in Lightweight CNN. This technique can extract additional discriminatory lesion characteristics and progress the performance in a bit of time. Next, two sets of output from CNN unit are used for training of two-feature group and differential networks simultaneously. Accuracy of 96.2% achieved by fusion performance can be improved by efficient discrimination network. Wang et al. [19] proposed an information that alert deep framework that imposes several clinical knowledge to feature segmentation and melanoma recognition unit. Lesion-based pooling and shape extraction (LPSE) structure is developed to move the data gained from image partition to detection unit. Also transmit data from recognition to segmentation units simultaneously. An efficient diagnosis guided feature fusion (DGFF) approach and recursive learning, is proposed which iteratively enhances the learning capability and boosts the performance It is more time consuming and labor intensive.

A lot of computer-aided diagnosis mechanisms have been created in the earlier. Due to the skin lesion images complicated visual qualities, which include irregularly shaped features and fuzzy edges, had trouble performing. Adegun and Viriri [20] suggested a deep learning technique to automate melanoma lesion detection and segmentation that gets over these restrictions. For efficient learning and feature extraction, an enhanced encoder-decoder structure with independent networks coupled through a number of skip pathways is proposed. This network increases the level of semantics of the feature encoder maps closest to decoder. In addition, SoftMax and Lesion-classifier were used. On International Skin on Biomedical Imaging (ISBI) 2017 dataset, proposed approach was accurate, with 95% and 92% dice coefficients. While on Pedro Hispano (PH2) dataset, it had an accuracy of 95%, dice coefficient of 93%. Some current state-of-the-arts are outperformed by this approach. This method needs better accuracy.

Manzo and Pellino [21] implemented deep CNN structures using prior training for visual representation with the aim of

predicting melanoma in skin lesions. To extract image features using a transfer learning strategy. The transfer learning features were then used in a grouping of classifications framework. The model precisely learns different classifiers then uses statistical techniques to grouping the supplied predictions. The primary weakness is the computing difficulty associated with the features extraction step, which is known to be time-consuming, especially as the amount of data to be analyzed increases. A hybrid classification method was created by İlkin et al. [22] combining a heuristic optimization technique and the SVM algorithm. It has certain advantages, but it also has some big disadvantages. Images with noise, such as hairs or colour changes, degrade performance. When a wounded area extends beyond the image's boundaries, the classifier's effectiveness in capturing the model's acquired characteristics diminishes.

### III. PROBLEM STATEMENT

When applying classification methods on Melanoma images, data imbalance was challenging. Data imbalance

causes overfitting problem [16], poor accuracy [20] and training error. Overfitting problem can be solved by increasing the size of dataset which accomplished through GAN (Generative Adversarial Network). CLAHE and Weiner filter which potentially remove noises and artifacts interfering with the classification. The existing work suffers from the issue of the gradient disappearing, which affects learning performance. To overcome these challenges this work proposes Bi-LSTM Deep Learning framework for skin cancer detection. The time consumption could be reduced by implementing African Vulture Optimization (AVO).

### IV. PROPOSED GAN BASED BI-LSTM WITH AVO

Research must automatically discover class-preserving changes to generate valid and representative samples in order to combat the imbalanced class problem and further improve classification accuracy. However, the samples of skin lesion should have high resolution for classification to detect the presence of malignant in skin lesion. Fig. 1 illustrates the proposed GAN-Bi-LSTM with AVO.

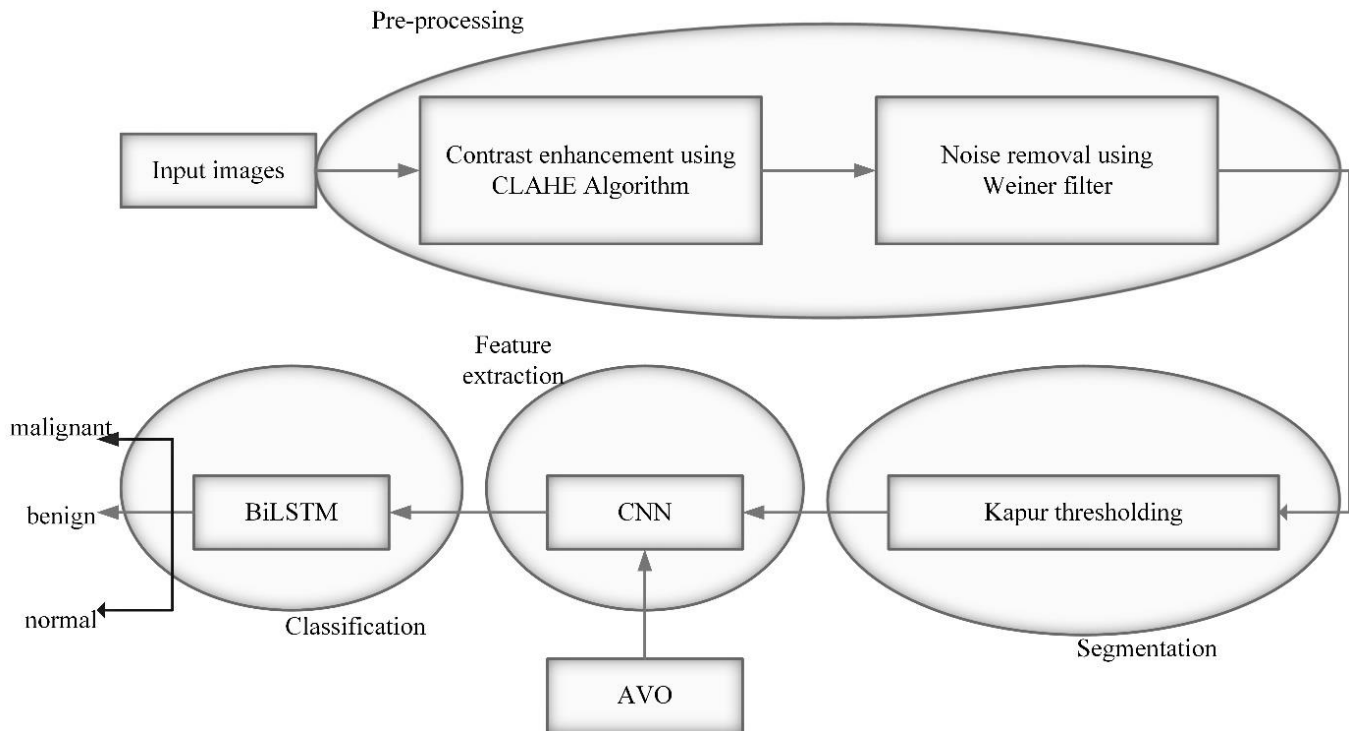


Fig. 1. Proposed AVO with GAN based Bi-LSTM.

The skin cancer images are grouped from ISIC 2018 and the fake images were generated using GAN to overpower the problem of data imbalance. The input images were preprocessed with CLAHE algorithm then filtered by Weiner filter to eliminate unwanted artifacts such as hair, veins and other noise factors. Then the skin contour is segmented by kapur threshold as binary images. The special features of skin images were extracted and optimum feature is selected in African Vulture Optimization (AVO). CNN with Bi-LSTM were designed to detect the melanoma from the skin lesion.

#### A. Data Collection

The International Skin Imaging Collaboration (ISIC 2018) provided the dataset for research which has challenges [23]. The objective is to automatically detect melanoma using dermoscopic pictures. There are three components to the challenge: Segmentation of Lesion, Visual Dermoscopic Features/Patterns Detection & Localization, and melanoma Classification. The proposed concentrate on the third job, which requires you to categorize lesion images into three different groups: melanoma, benign and normal. The dataset contains of images with multiple resolution. Although researcher chooses the test images at random, make sure that the class distribution is maintained and that the proportion of images from each class is comparable in the train and test sets.



B. GAN-based Data Augmentation

GAN is one of the generative techniques in deep learning. GAN framework consists of two neural networks. They are simultaneously guided discriminator and generator. A generator network G produces images from noise n and given to discriminator network D which identify the difference

between real image y and synthetic images generated by generator G. The images with better resolution are ready for further process. If the image produced by generator is not identical to the real image again it is given to discriminator and generator network. The construction of GAN is depicted in Fig. 2.

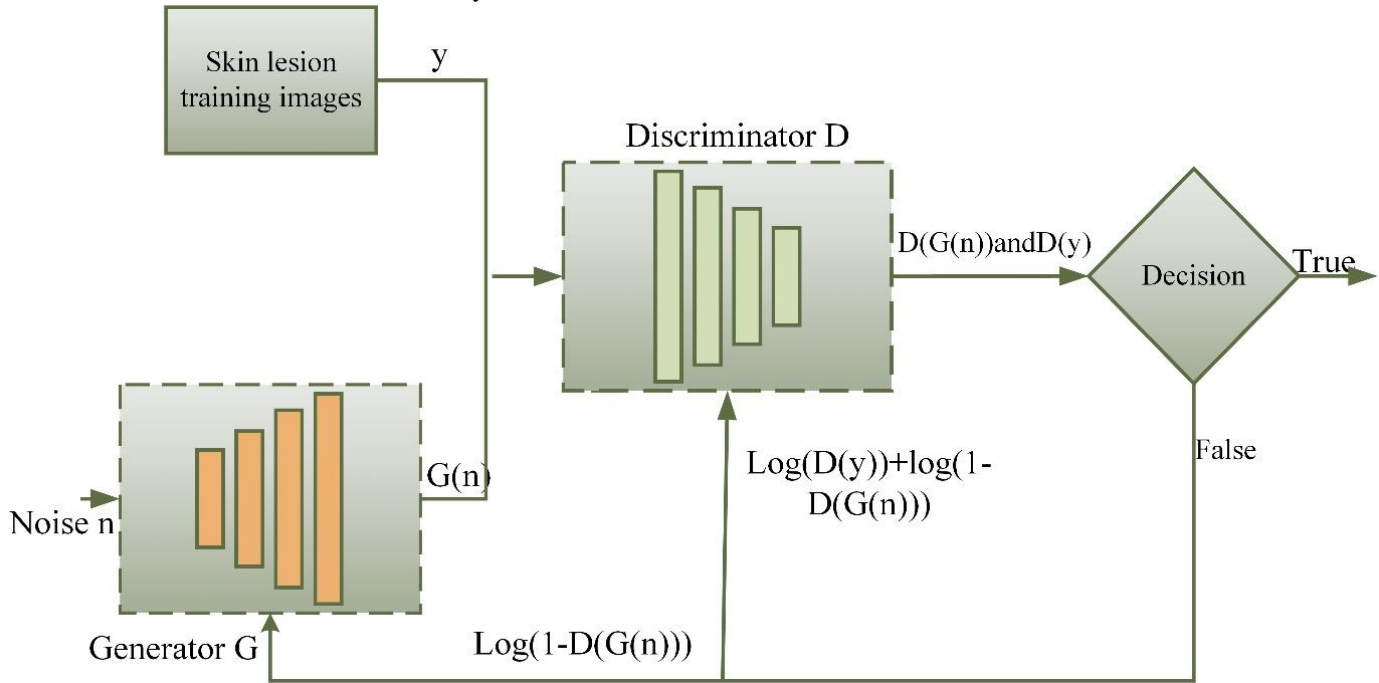


Fig. 2. Construction of GAN.

1) *Generator*: A convolution neural network serves as the generator based on unsupervised learning. The generative model that creates images that is comparable to the training images. In order to produce samples with a distribution that is similar to that of real samples, this artificial neural network assesses the probable distribution of the raw data and updates its parameters. The generator gets noise that is random and provides outputs information. There is some noise in the distribution as well. Now, the distribution of noise can be normal, uniform, or any other type. The generator feeds the discriminator network with its fake image output, and the discriminator performs its training and determines whether the input is real or fake.

2) *Discriminator*: A convolution neural network is a discriminator with supervised learning. It works as binary classifier that is learned on training images and forecasts whether the test image is a genuine or one that has been manufactured. Real images from the original training samples as well as generated image produced by the generator are included in the discriminator's training samples. During the training process, these fake images are utilised as negative examples. The generator works effectively creating genuine images. Only the discriminator can fail to discriminate among the first and generated samples.

3) *Loss function*: The loss function is found to evaluate error then it restores variations. It is specified for the generator

as well as the discriminator. The difference among the delivery of data generated by the generator and the delivery of the real images is calculated using the GAN loss function [24]. The discriminator tries to lower the negative log-likelihood. However, the generator just maximises negative log-likelihood since it wants to trick the discriminator. Since it can use the discriminator's cost function, the discriminator loss function is given in Eq. (1) as:

$$\max_D V(G, D) = E_{y \sim P_{data}(y)} [\log D(y)] + E_{y \sim P_n(n)} [\log(1 - D(G(n)))] \quad (1)$$

To minimize  $\log(1 - D(G(n)))$ , the Generator is trained. It produces images as similar to the real training images as possible. The generator loss function is expressed in Eq. (2) as,

$$\min_G V(G, D) = E_{y \sim P_n(n)} [\log(1 - D(G(n)))] \quad (2)$$

The generator and discriminator are optimized by the following Eq. (3),

$$\min_G \max_D V(D, G) = E_{y \sim P_{data}(y)} [\log D(y)] + E_{n \sim P_n(n)} [\log(D(G(n)))] \quad (3)$$

C. Image Preprocessing

Progressed images with interrelated masks for rotation, resizing, reflection and brightness were designed for each image. The less quality of the input lesion images delivered by electronic detectors restricts detection and evaluation. Up

sampling was used on the lesion images to solve the imbalanced class distribution. The ISIC2018 dataset was partitioned into three distinct groups for training, testing, and analysis to deal with the overfitting problem carried on by the smaller quantity of used training images.

1) *Image enhancement using contrast-limited adaptive histogram equalization (CLAHE)*: The pixel spreading can be realised in the image histogram. The contrast of image can be improved by rearranging the pixel distribution. Histogram equalisation, which can improve the variety of every pixel grey amount, is a mapping transformation of the original image 's grey level. So that image contrast is enlarged. An adaptive histogram equalisation (AHE) technique has a

tendency to overamplify noise in the areas of the image that are reasonably uniform. The CLAHE approach was suggested as a solution to this issue. Divide the image into parts that are non-overlap. Typically, the area dimension is established to 8 by 8. Obtain the histogram for each region, then trim the histogram using the threshold. By clipping of histogram with a predetermined threshold prior to calculating the Cumulative Distribution Function (CDF), the CLAHE algorithm attains the purpose of restricting the amplification [25]. This restricts the transformation function's slope as well. Redistribute pixels, then uniformly Spread the values of the clipped pixels under the histogram. local histogram equalisation is carried out on each region is shown in Fig. 3.

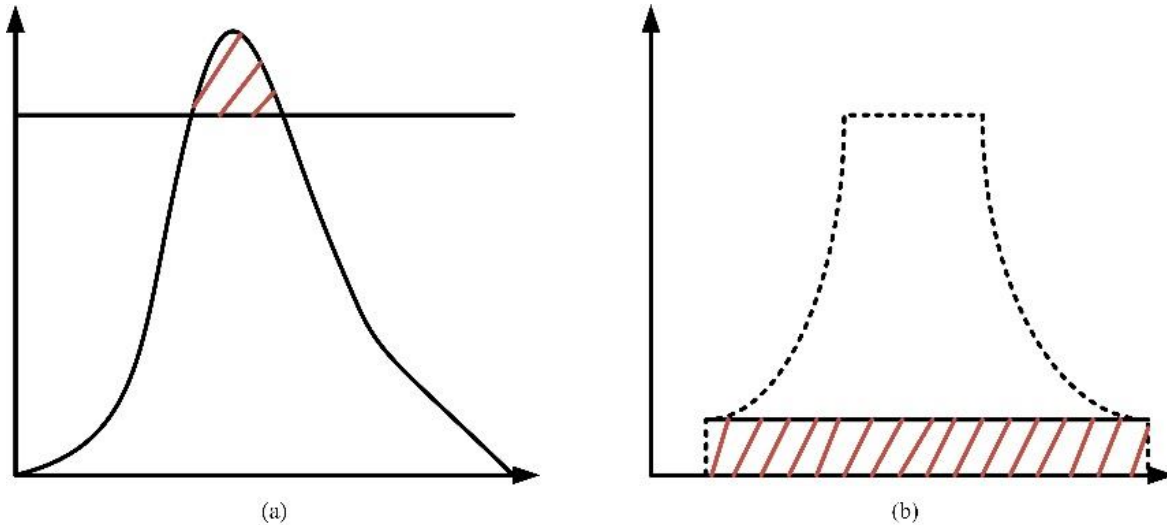


Fig. 3. Histogram equalization (a) before cutting (b) after cutting.

The linear interpolation is used to reconstruct the pixel value. Consider  $v$  as a grey value of image's sample point  $R$  and  $v'$  is the new grey value of that by performing linear interpolation. Let the sample points for surrounding regions are  $R_1, R_2, R_3$  and  $R_4$ .  $gr(v)$  is grey-level mapping for  $v$ .

The new grey value for pixels in the corners corresponds to the grey-level mapping for  $v$ . The new grey

value is expressed in Eq. (4),

$$v' = gr_1(v) \tag{4}$$

New grey value of pixels in edges is the mapping of grey level for  $v$  of two samples is expressed in Eq. (5),

$$v' = (1 - \alpha)gr_1(v) + \alpha gr_2(v) \tag{5}$$

New grey value of pixels in centre is the mapping of grey level for  $v$  of four samples is given in Eq. (6),

$$v' = (1 - \beta)((1 - \alpha)gr_1(v) + \alpha gr_2(v)) + \beta((1 - \alpha)gr_3(v) + \alpha gr_4(v)) \tag{6}$$

where, the normalized distances are  $\alpha$  and  $\beta$  with regards to the point  $R_1$ .

Due to the reason that a few of the images have tiny pixel counts and which need to be resized. This leads to

considerable changes in the image's luminance and size. There are many sets of parameters for various acquisition instruments. All pixel density was normalized within the range  $[-1, 1]$  to ensure that the data were reliable and noise-free. Eq. (7)'s normalization computation made the model less sensitive to minute weight changes. Normalization of image  $I_{Norm}$  is given in Eq. (7) as,

$$I_{Norm} = (I - min_i) \left( \frac{2}{max_i - min_i} \right) - 1 \tag{7}$$

Where,  $min_i$  and  $max_i$  are minimum image and maximum image.

2) *Noise removal by wiener filters*: The method used to remove unwanted data from the image is statistical. It achieves ideal trading among noise flattening and reverse filtering which filter the blurring and noise existing in the image [26].

Filter function is given in Eq. (8) as,

$$f(y, z) = \left[ \frac{H(y, z)^*}{H(y, z)^2 + \frac{S_n(y, z)}{S_i(y, z)}} \right] G(y, z) \tag{8}$$

Where  $G(y, z)$  represents degraded image,  $H(y, z)$  is degradation function,  $S_n(y, z)$  is a power spectra of noise and  $S_i(y, z)$  shows the original image's power spectrum.

#### D. Image Segmentation by Kapur Thresholding Technique

When taking a picture in many medical circumstances, negative impacts imposed to the skin image, causing image examination challenging for the computational approaches. In the case of input, portion of the skin image is essential and the rest of the image is not significant. To eliminate these negative consequences, kapur thresholding was applied. The image thresholding is used to keep the essential parts and eliminate the unwanted. Image thresholding is an image separation method that converts an image's pixel values to zero and one. In general, a threshold level value for the picture pixel points which overall value as threshold for all or each pixel may have different threshold value. The image's pixels are then compared to that level, and if the pixel intensity is more than the threshold, it is turned white; otherwise, it is turned black. As an outcome, a grayscale image is transformed into an image that is binary (black and white). Typically apply thresholding to choose portions of an image and eliminate those that aren't significant to us. Consider an image with L levels and N pixels ranging between 0 and L-1, If  $g(j)$  indicates the gray-level number and j describes image existences then median value of image is given in Eq. (9).

$$Mean = g(j)/N \quad (9)$$

The dermoscopy pictures were threshold using the Kapur technique. This procedure determines the limit using entropy  $T(u)$  boosting and histogram data, which can be computed using Eq. (10) to (14):

$$Max T(u) = A(0, u) + A(u, L) \quad (10)$$

$$Where \quad A(0, u) = - \sum_{j=0}^{u-1} \frac{P_j}{W_0} \ln \frac{P_j}{W_0} \quad (11)$$

$$W_0 = \sum_{j=0}^{u-1} P_j \quad (12)$$

$$A(u, L) = - \sum_{j=u}^{L-1} \frac{P_j}{W_1} \ln \frac{P_j}{W_1} \quad (13)$$

$$W_1 = \sum_{j=u}^{L-1} P_j \quad (14)$$

The ideal threshold value is calculated by maximizing the function  $T(u)$  and u number of grey levels was considered.

#### E. Extraction and Classification using AVO based CNN -Bi-LSTM

1) *Feature extraction using CNN*: Convolution, fully connected (FC) and pooling layers are stacked to form a CNN architecture. Each convolution layer has a set of filters that are learnable which is to acquire local features from the image being processed using the knowledgeable filters. It is probable to reduce computing complication while improving performance by constructing filters that conduct convolution actions based on two important ideas, namely weight division and local linking. The pooling layer is in charge of the down sampling process. One of the pooling layer's distinguishing characteristics is that it reduces the overall dimension of the image and avoids overfitting. Typically, FC layers are employed in the final CNN layer design to learn features returned by the layer of convolution; it is then utilised to construct the output.

2) *Feature selection using African vulture optimization (AVO)*: The latest metaheuristic technique developed through vulture tracking is called AVO. Vultures are a type of bird that may be found all over the world. Vultures are typically carnivorous; nevertheless, these birds are unable to divide the flesh and must rely on a different meat-eating to do it. African vultures may reach altitudes that exceed 11000 meters, and they travel great distances and rotate to locate food. They usually have difficulties after discovering a food source. The weaker vultures encircle the healthier vultures, delaying their activity; as that vulture fatigue, they begin to seek food. This way of living motivates to create a novel metaheuristic approach for tackling an optimization issue [27].

**Stage 1:** Increasing the size of the population, achieving similar values for each vulture, identifying the greatest vulture in each group, and selecting the best result for each group. This is explained in Eq. (15).

$$S(k) = \begin{cases} Best \ vulture_1, & if \ Z_k = f_1 \\ Best \ vulture_2, & if \ Z_k = f_2 \end{cases} \quad (15)$$

$f_1$  and  $f_2$  specify the parameters that are examined before to optimization, which must be between 0 and 1, where  $f_1+f_2=1$ .

**Stage 2:** Calculating the vulture famine rate. Vultures soar over the sky in quest of food. When there is insufficient energy, the vulture approaches stronger to grab a free meal. This is mathematically modelled in Eq. (16) and (17) as,

$$V = (2 \times \delta + 1) \times N \times \left(1 - \frac{iter_k}{max_{iter}}\right) + l \quad (16)$$

Where,

$$l = d \times \left(\sin\left(\frac{\pi}{2} \times \frac{iter_k}{max_{iter}}\right) + \cos\left(\frac{\pi}{2} \times \frac{iter_k}{max_{iter}}\right) - 1\right) \quad (17)$$

Where  $\delta$  denotes a random value ranged between [0,1],  $iter_k$  denotes the current iteration, N denotes a fixed number which displays the procedure of the optimization and makes the investigation and process stages,  $max_{iter}$  denotes the overall number of iterations, and d indicates an amount that is limited between -2 and 2. If N falls below zero, the vulture is hungry, and if it rises over one, the vulture is fulfilled.

**Stage3:** Enquiry.

In this method vultures contain random segments with two potential designs and a variable  $R_1$  with a value ranging from zero to one to determine the plane. The mathematical procedure for finding a meal for vultures is given in Eq. (18) and (19) as follows:

If  $R_1 \geq randR_1$

$$S(k+1) = BV(k) - T(k) \times S \quad (18)$$

If  $R_1 < randR_1$

$$S(k+1) = BV(k) - S + rand_2 \times ((ub - lb) \times rand_3 + lb) \quad (19)$$

where S identifies the vultures that diverge from others at randomly in search of food., lb and ub are the variables' lower

and upper bounds,  $BV$  contains the best vultures, and  $rand_2$  and  $rand_3$  define two random standards among 0 and 1.

**Stage 4: Utilization.**

This happens if  $|S| < 1$  divides the phase into two halves with two strategies that are determined by two R2 and R3 constraints that are in the range  $[0, 1]$ . The first portion of utilization begins at  $0.5 < |S| < 1$ . Both designs include rotational flights. If  $|S| \geq 0.5$  indicates that a vulture is spirited; in this stage, weak bird attempt to ingest food on stronger vultures.  $S(k + 1)$  represents the vulture's present location and can be designed using Eq. (20), (21), (22) as follows:

$$S(k + 1) = \frac{B_1 + B_2}{2} \quad (20)$$

Where,

$$B_1 = Best\ vulture_1(k) - \frac{Best\ vulture_1(k) \times S(k)}{Best\ vulture_1(k) - S(k)^2} \times S \quad (21)$$

$$B_2 = Best\ vulture_2(k) - \frac{Best\ vulture_2(k) \times S(k)}{Best\ vulture_2(k) - S(k)^2} \times S \quad (22)$$

The initial step of optimization is to specify the algorithm's minimal and maximum rate boundaries for avoiding system failures. The least acceptable value for the maximum pooling

is taken to be 2, and the highest value is considered as moving width. The proposed CNN's half-value accuracy has been used as an objective function in this case. The algorithm was initialized, updated, and reached the end state before the procedure was ended. Weights and biases, which make up a CNN's main building blocks, are thought to be enhanced.

3) *Skin cancer classification through Bi-LSTM:* Bidirectional LSTM layers extract hidden and sequential features from images in both forward and reverse time directions. RNN involves memory and data storage limitations. It is incapable of learning long-term, which might lead to gradient disappearance. As a result, to overpower the insufficiencies of algorithm, the LSTM approach was devised. This construction is formed on the usage of memory cells which store long-term and gate mechanisms to control this information. A basic LSTM unit contains three types of gates: Input gate  $i_t$ , Forget gate  $m_t$ , and output gate  $o_t$ . Each gate controls the state of memory cells by executing point-wise multiplication and sigmoid operations on the image  $y_t$ . The architecture of LSTM is exposed in Fig. 4.

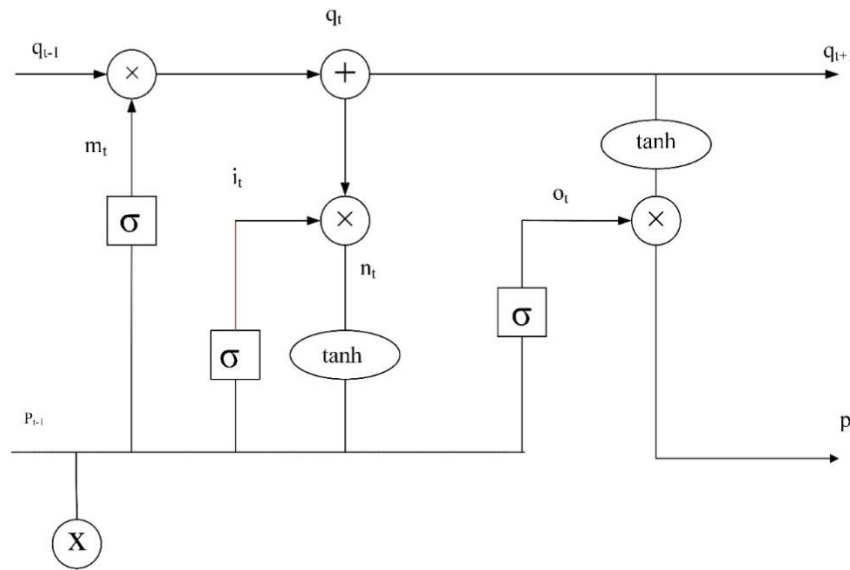


Fig. 4. LSTM architecture.

When input data  $y_t$  at its present state and output  $p_{t-1}$  from the preceding layer's hidden state are both entered, all gates are triggered. A forget gate specifies what information should be retained and which data should be ignored. The data gathered from the present input is used by sigmoid functions  $y_t$  to convey information from the present input  $y_t$  to the previous hidden state  $p_{t-1}$ . The forget gate's output value is between 0 and 1. When the rate is near to zero that indicates the data will be removed. More knowledge tends to be kept closer to oneself. The following procedures must be followed to calculate the forget gate formula using Eq. (23) input gate formula using Eq. (24) and output gate formula using Eq. (27) as follows:

$$m_t = \sigma(w_m \cdot [p_{t-1}, y_t] + b_m) \quad (23)$$

$$i_t = \sigma(w_i \cdot [p_{t-1}, y_t] + b_i) \quad (24)$$

$$\hat{q}_t = \tanh(w_q \cdot [p_{t-1}, y_t] + b_q) \quad (25)$$

Current state equation is expressed in eqn. (26),

$$q_t = m_t \odot q_{t-1} + i_t \odot \hat{q}_t \quad (26)$$

$$o_t = \sigma(w_o \cdot [p_{t-1}, y_t] + b_o) \quad (27)$$

Hidden state formula is given in eqn. (28),

$$p_t = o_t \odot \tanh(q_t) \quad (28)$$

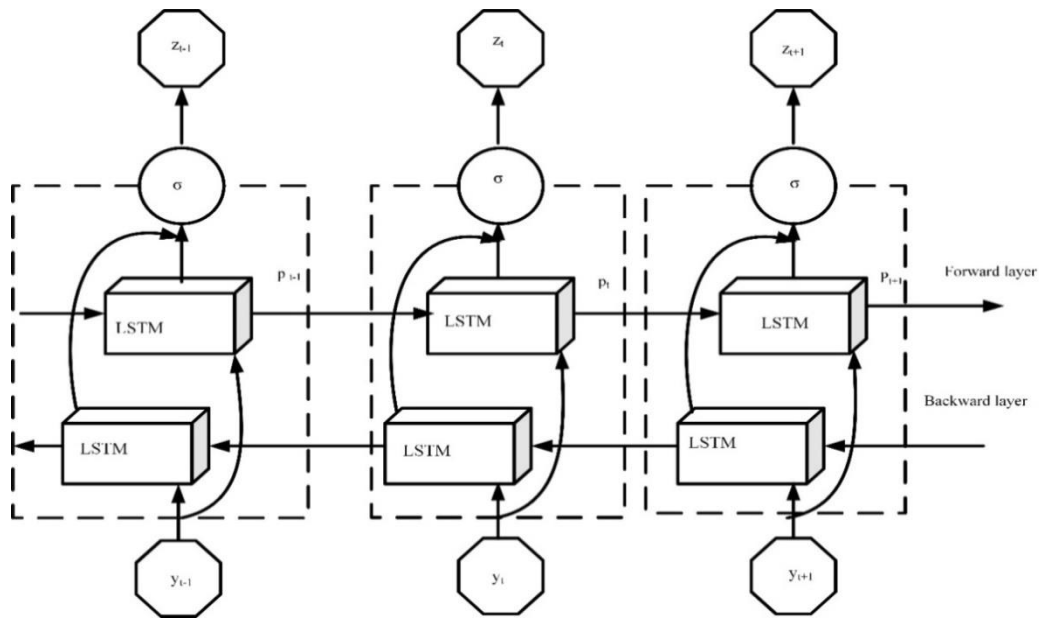


Fig. 5. Bi-LSTM architecture.

Bidirectional LSTM (Bi-LSTM) links the two LSTM hidden layers to the output layer. Combining two LSTM as a single layer stimulates enhance the learning long-term dependence and model performance. Fig. 5 depicts the

structure of an unfolded Bi-LSTM layer with LSTM layer in forward and a backward direction. It classifies the input images as melanoma, benign and normal. Thus, the melanoma is detected from the input images.

---

**GAN-AVO Algorithm**

---

*Input: Image containing skin lesion*

*Output: Melanoma detection in skin image*

*Load the input image*

*Image augmentation using GAN*

*Perform preprocessing operation*

*//contrast enhancement using CLAHE*

*Noise removal using Wiener filter is given in eqn. (8)*

*Image segmentation*

*//compute threshold value using eqn. (10)*

*Feature extraction*

*//color, border, diameter, shapes were extracted using CNN*

*Select feature using AVO*

*Calculate the fitness of vultures*

*If ( $|s| \geq 1$ ) then*

*Upgrade the location vulture using eqn. (19)*

*Else*

*Upgrade the location vulture using eqn. (20)*

*Endif*

*Classification using Bi-LSTM*

*End*

---

The GAN-AVO algorithm is designed for the discovery of melanoma in skin lesion images. It begins by taking an input image containing a skin lesion and applies image

augmentation through a Generative Adversarial Network (GAN) for data enhancement. Subsequently, the image undergoes preprocessing, involving contrast enhancement

using the CLAHE method and noise removal through a Weiner filter. Image segmentation is performed by computing a threshold value. To retrieve relevant features, a Convolutional Neural Network (CNN) is employed to capture color, border, diameter, and shape information. AVO is utilized to select pertinent features, followed by the computation of fitness for vultures. If the number of selected features ( $|s|$ ) is greater than or equal to 1, the algorithm

upgrades vulture locations using Eq. (19); otherwise, it utilizes Eq. (20). Finally, the algorithm employs a Bi-LSTM network for classification, culminating in melanoma detection. This comprehensive approach integrates GAN-based augmentation, preprocessing, segmentation, feature extraction, AVO feature selection, and Bi-LSTM classification to effectively identify melanoma in skin lesion images. Fig. 6 shows flowchart of this research.

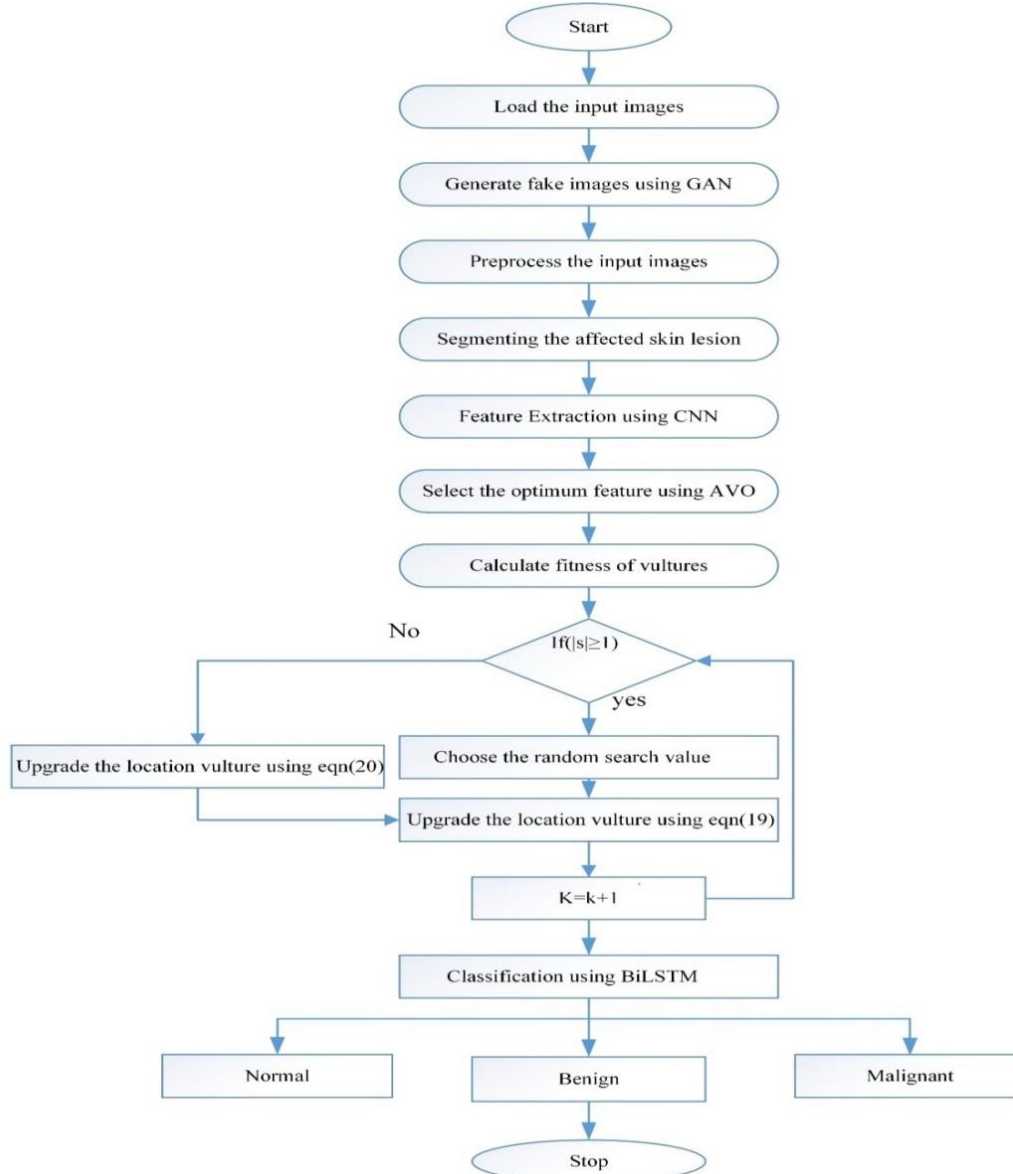


Fig. 6. Flowchart of proposed method.

## V. RESULTS

The proposed study aimed to enhance the accuracy of melanoma detection in skin images by employing a novel AI driven African Vulture optimization with GAN based Bi-LSTM deep framework for melanoma detection. The approach used GANs to create realistic and variety fake images, which were then mixed with the original dataset to supplement training images and improve model generalisation. AVO was used to optimise the network design and hyperparameters,

improving the model's overall performance. Extensive tests on a large-scale image were carried out, and the findings showed a considerable improvement in skin cancer diagnosis and localization accuracy. The following metric was used to assess the model efficiency of the strategy.

For comparison, the following segmentation of skin lesion evaluation criteria was used: precision, recall, F1-score, and accuracy. These parameters were used to assess the model. These are depicted below:

**Accuracy:** It calculates the fraction of real outcomes including true positives and true negatives across all cases investigated. It is expressed in Eq. (29),

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (29)$$

**Precision:** The ratio of exactly anticipated positive outcomes to overall predicted positive occurrences is defined as precision. The precision is calculated using Eq. (30).

$$Precision = \frac{TP}{TP+FP} \quad (30)$$

**Recall:** The recall measures the proportion of genuine positive samples that were projected to be positive. Using Eq. (31), calculate the value recall.

$$Recall = \frac{TP}{TP+FN} \quad (31)$$

where, FP represents false positive pixels, FN signifies false negative pixels, TP symbolizes true positive pixels, and TN describes true negative pixels.

**F1-score:** In the categorization task, recall and accuracy relate to one another. Although a high value for both is ideal, the reality is generally great accuracy with low recall, or high recall with low accuracy. To account for both recollection and accuracy, the F1-score, which is a mean of recall and accuracy, can be employed. Eq. (32) shows the definition of F1-score.

$$F1 - score = 2 * \frac{Precision*Recall}{Precision+Recall} \quad (32)$$

TABLE I. PERFORMANCE METRICS OF PROPOSED METHOD

Proposed GAN-BiLSTM with AVO	
Performance Metrics	Values (%)
Accuracy	98.5
Precision	98.1
Recall	98
F1-score	98

The assessment results of the created skin cancer detection system employing the combined strategy are shown in Table I. At 98.5%, the accuracy is exceptionally high. Precision, which measures the percentage of accurate positive predictions compared to all positive forecasts, is an outstanding 98.1%. The system's capacity to accurately detect real positive cases is demonstrated by the recall measure, also known as sensitivity which is remarkably high at 98%. At 98%, the F1-score, which balances recall and accuracy, is very impressive.

With high values across key performance parameters, these findings show precise and reliable skin cancer diagnosis. Fig. 7 illustrates the performance assessment of proposed GAN-Bi-LSTM with AVO.

Table II shows the Accuracy, Recall Precision and F1-score of the proposed approach with existing methods. The accuracy of the suggested method GAN-Bi-LSTM with AVO (98.5%) is higher than the existing approaches convnet (91.03%), Inception RESNET (91%) and RESNET-18(94.47%). Fig. 8 depicts the graphic depiction of the

performance metrics of proposed with existing approaches. The precision of the suggested method GAN-Bi-LSTM with AVO (98.1%) is higher than the existing approaches convnet (91.09%), Inception RESNET (91%) and RESNET-18(93.57%). The recall of the suggested method GAN-Bi-LSTM with AVO (98%) is higher than the existing approaches convnet (90.96%), Inception RESNET (91%) and RESNET-18(94.01%). The F1-score of the suggested method GAN-Bi-LSTM with AVO (98%) is higher than the existing approaches convnet (91.01%), Inception RESNET (91%) and RESNET-18(94.45%).

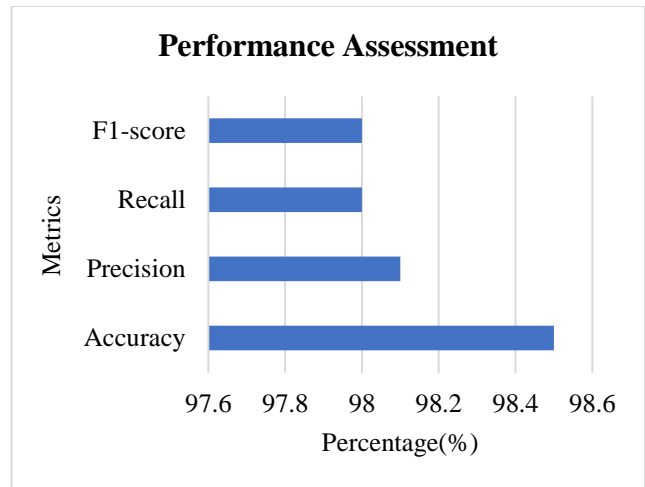


Fig. 7. Performance assessment of proposed GAN-Bi-LSTM with AVO.

TABLE II. PERFORMANCE METRICS OF PROPOSED METHOD IS EVALUATED WITH EXISTING METHODS

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Convnet [28]	91.03	91.09	90.96	91.01
Inception RESNET [29]	91	91	91	91
RESNET-18 [30]	94.47	93.57	94.01	94.45
Proposed method	98.5	98.1	98	98

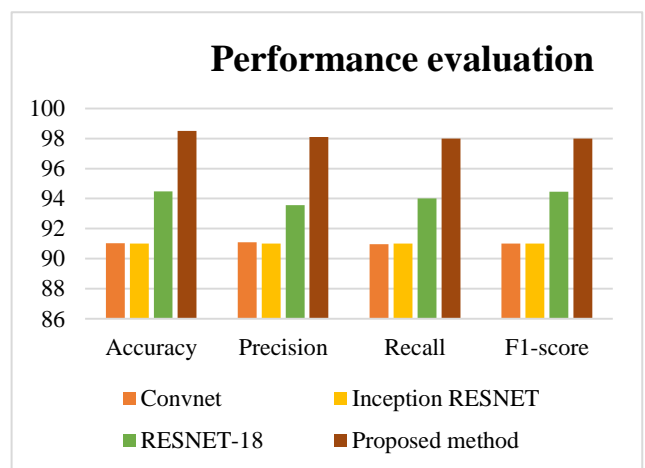


Fig. 8. Graphical illustration of the performance metrics of proposed with existing approaches.

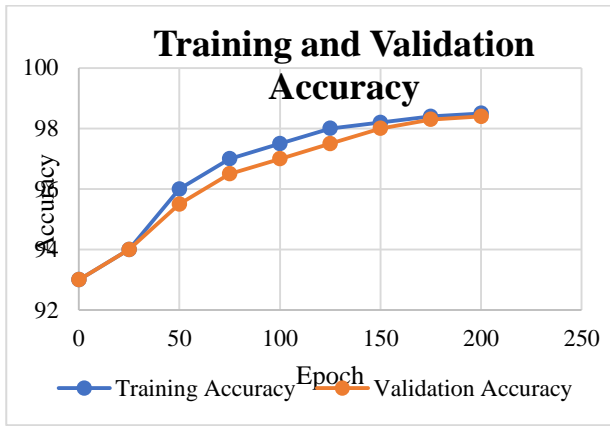


Fig. 9. Graphical depiction for training and validation accuracy of proposed method.

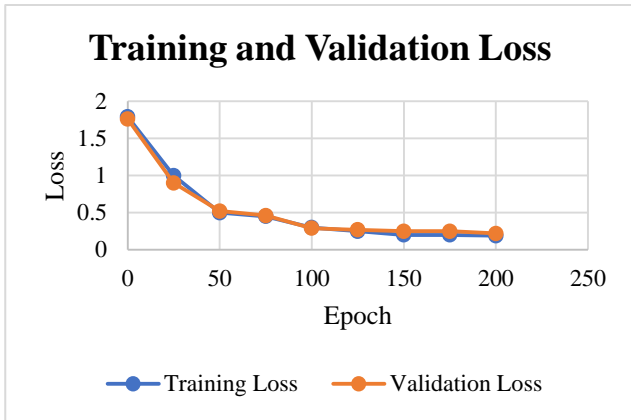


Fig. 10. Graphical representation of loss in proposed AVO-Bi-LSTM.

The accuracy level and loss rates fluctuation graphs for the entire GAN-based AVO-Bi-LSTM model procedure are displayed in Fig. 9 and 10. The accuracy ratio and loss ratio overall graph has stabilized at the training intervals of the GAN-based AVO-Bi-LSTM are at 100, and it is clear that the GAN-based AVO-Bi-LSTM fits data more quickly.

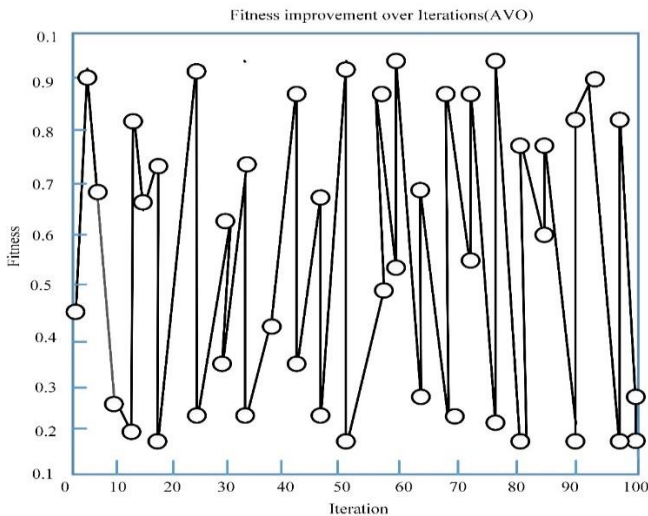


Fig. 11. Fitness improvement over iterations (AVO).

Before optimisation the value of accuracy for proposed GAN based AVO-Bi-LSTM is 98%. The accuracy achieved after optimisation using AVO is 98.5%. The fitness of AVO is depicted in Fig. 11. Using Eq. (33), the Area Under the Curve (AUC) has been calculated to estimate AVO-Bi-LSTM's overall performance. Fig. 12 shows ROC curve for the GAN-based AVO-Bi-LSTM model, and it can be seen that the ROC area is nearly close to 1, confirming the model's good stability and potential for usage as classification model for skin cancer diagnosis.

$$AUC = \frac{1}{2} \left( \frac{TP}{TP+FN} + \frac{TN}{TN+FP} \right) \quad (33)$$

AUC of proposed GAN based AVO-Bi-LSTM is compared with existing model (mAlexNet + Bi-LSTM) is given in Table III.

TABLE III. AUC OF PROPOSED GAN BASED AVO-Bi-LSTM IS COMPARED WITH EXISTING MODEL

Architecture	AUC
mAlexNet + Bi-LSTM	0.97
Proposed AVO + Bi-LSTM	0.985

In Table III, two distinct architectures are evaluated for their melanoma detection capabilities. The first architecture, which combines modified AlexNet (mAlexNet) with a Bi-LSTM network, achieves an Area Under the Curve (AUC) of 0.97. This indicates a strong ability of the mAlexNet and Bi-LSTM combination to victimize melanoma. In comparison, the proposed approach, utilizing Artificial Vulture Optimization (AVO) for feature selection in conjunction with a Bi-LSTM network, outperforms the former with an AUC of 0.985. The higher AUC value achieved by the proposed AVO + Bi-LSTM architecture underscores its superior ability to effectively distinguish melanoma cases from non-melanoma instances, highlighting its potential as an advanced and promising model for accurate melanoma detection in skin lesion images.

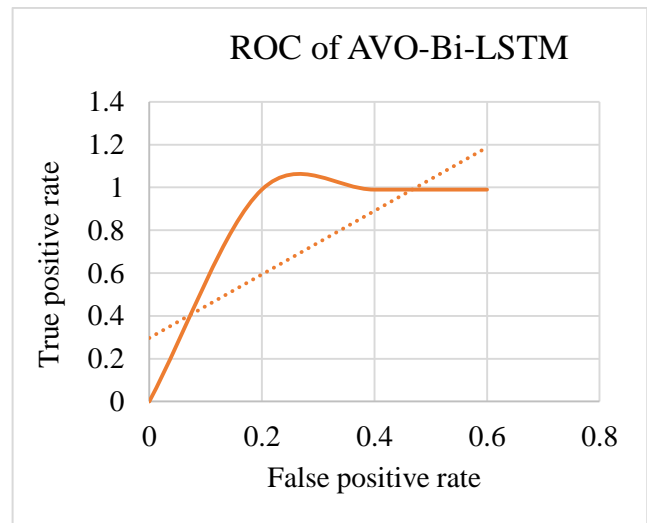


Fig. 12. ROC curve for proposed AVO-Bi-LSTM.



## A. Discussion

The proposed study aimed to significantly enhance the accuracy of melanoma detection in skin images by employing a novel AI-driven approach combining African Vulture optimization (AVO) with a Generative Adversarial Network (GAN)-based Bidirectional Long Short-Term Memory (Bi-LSTM) deep framework. This innovative approach harnessed GANs to generate realistic and diverse synthetic images, augmenting the original dataset to improve model generalization. AVO was utilized to optimize network design and hyperparameters, resulting in a substantial boost in the overall performance of the model. Extensive testing on a large-scale image dataset yielded highly promising results, showcasing a remarkable improvement in skin cancer diagnosis and localization accuracy. The evaluation of the model's efficiency was based on standard segmentation criteria, including precision, recall, F1-score, and accuracy. The results were exceptionally favorable, with an accuracy rate of 98.5%, precision at 98.1%, recall reaching 98%, and an impressive F1-score of 98%, indicating the model's precision and reliability in skin cancer diagnosis.

Comparing the proposed approach with existing methods, the superiority of the GAN-Bi-LSTM with AVO method was evident. Its accuracy of 98.5% outperformed other methods such as Convnet (91.03%), Inception RESNET (91%), and RESNET-18 (94.47%). Similarly, the precision, recall, and F1-score of the proposed method were significantly higher than those of existing methods, reaffirming its effectiveness in accurate melanoma detection. Visual representations of performance metrics, as depicted in Fig. 7, 8, 9, and 10, further illustrated the consistent and stable performance of the GAN-Bi-LSTM with AVO model. Moreover, AVO optimization yielded an even higher accuracy of 98.5%, compared to the initial accuracy of 98%, as shown in Fig. 11. The Area Under the Curve (AUC) analysis also indicated the superiority of the proposed AVO + Bi-LSTM architecture over an existing model, with an AUC of 0.985 compared to 0.97, underscoring its potential as an advanced and reliable model for accurate melanoma detection in skin lesion images.

## VI. CONCLUSION

The research introduces an AI-powered framework that combines African Vulture Optimization with Generative Adversarial Networks (GANs) and Bidirectional Long Short-Term Memory (Bi-LSTM) networks for melanoma detection. This study serves as an illustrative example of Bi-LSTM's application in identifying skin cancer from lesion images. GANs are harnessed to address data imbalance, generating additional data to enhance detection. Furthermore, the research devises an ensemble model that synergizes Bi-LSTM and GANs, surpassing the performance of the standalone deep learning model. Employing an iterative AVO evolutionary method enhances a population of solutions via fitness assessment. Depending on outcomes, this approach could potentially outperform existing artificial intelligence methods. Ultimately, the proposed hybrid deep learning framework, incorporating GANs and AV optimization, notably improves the accuracy of skin cancer detection and localization within skin images. The integration of GANs and AV optimization

effectively tackles data limitations and optimizes deep learning models, contributing to the advancement of melanoma diagnosis. The future work explores the integration of other imaging modalities, such as dermoscopy or infrared imaging, alongside the proposed framework to enhance the model's ability to detect melanoma across various imaging sources.

## REFERENCES

- [1] R. Ashraf et al., "Region-of-Interest Based Transfer Learning Assisted Framework for Skin Cancer Detection," IEEE Access, vol. 8, pp. 147858–147871, 2020, doi: 10.1109/ACCESS.2020.3014701.
- [2] M. S. Khan, K. N. Alam, A. R. Dhruba, H. Zunair, and N. Mohammed, "Knowledge Distillation approach towards Melanoma Detection," Comput. Biol. Med., vol. 146, p. 105581, Jul. 2022, doi: 10.1016/j.combiomed.2022.105581.
- [3] D. Adla, G. V. R. Reddy, P. Nayak, and G. Karuna, "Deep learning-based computer aided diagnosis model for skin cancer detection and classification," Distrib. Parallel Databases, vol. 40, no. 4, pp. 717–736, Dec. 2022, doi: 10.1007/s10619-021-07360-z.
- [4] A. Adegun and S. Viriri, "Deep learning techniques for skin lesion analysis and melanoma cancer detection: a survey of state-of-the-art," Artif. Intell. Rev., vol. 54, no. 2, pp. 811–841, Feb. 2021, doi: 10.1007/s10462-020-09865-y.
- [5] M. Deb Barma, M. A. Indiran, P. Kumar R, A. Balasubramaniam, and M. P. S. Kumar, "Quality of life among head and neck cancer treated patients in South India: A cross-sectional study," J. Oral Biol. Craniofacial Res., vol. 11, no. 2, pp. 215–218, Apr. 2021, doi: 10.1016/j.jobcr.2021.02.002.
- [6] K. Thurnhofer-Hemsi and E. Domínguez, "A Convolutional Neural Network Framework for Accurate Skin Cancer Detection," Neural Process. Lett., vol. 53, no. 5, pp. 3073–3093, Oct. 2021, doi: 10.1007/s11063-020-10364-y.
- [7] S. Banerjee, S. K. Singh, A. Chakraborty, A. Das, and R. Bag, "Melanoma Diagnosis Using Deep Learning and Fuzzy Logic," Diagnostics, vol. 10, no. 8, p. 577, Aug. 2020, doi: 10.3390/diagnostics10080577.
- [8] A. Naeem, M. S. Farooq, A. Khelifi, and A. Abid, "Malignant Melanoma Classification Using Deep Learning: Datasets, Performance Measurements, Challenges and Opportunities," IEEE Access, vol. 8, pp. 110575–110597, 2020, doi: 10.1109/ACCESS.2020.3001507.
- [9] R. M. Abd El-Aziz et al., "An Effective Data Science Technique for IoT-Assisted Healthcare Monitoring System with a Rapid Adoption of Cloud Computing," Comput. Intell. Neurosci., vol. 2022, pp. 1–9, Jan. 2022, doi: 10.1155/2022/7425846.
- [10] L. Ichim and D. Popescu, "Melanoma Detection Using an Objective System Based on Multiple Connected Neural Networks," IEEE Access, vol. 8, pp. 179189–179202, 2020, doi: 10.1109/ACCESS.2020.3028248.
- [11] J. K. Winkler et al., "Melanoma recognition by a deep learning convolutional neural network—Performance in different melanoma subtypes and localisations," Eur. J. Cancer, vol. 127, pp. 21–29, Mar. 2020, doi: 10.1016/j.ejca.2019.11.020.
- [12] M. Nawaz et al., "Skin cancer detection from dermoscopic images using deep learning and fuzzy k -means clustering," Microsc. Res. Tech., vol. 85, no. 1, pp. 339–351, Jan. 2022, doi: 10.1002/jemt.23908.
- [13] M. Toğaçar, Z. Cömert, and B. Ergen, "Intelligent skin cancer detection applying autoencoder, MobileNetV2 and spiking neural networks," Chaos Solitons Fractals, vol. 144, p. 110714, Mar. 2021, doi: 10.1016/j.chaos.2021.110714.
- [14] H. U. Rehman, N. Nida, S. A. Shah, W. Ahmad, M. I. Faizi, and S. M. Anwar, "Automatic melanoma detection and segmentation in dermoscopy images using deep RetinaNet and conditional random fields," Multimed. Tools Appl., vol. 81, no. 18, pp. 25765–25785, Jul. 2022, doi: 10.1007/s11042-022-12460-8.
- [15] B. A. Albert, "Deep Learning From Limited Training Data: Novel Segmentation and Ensemble Algorithms Applied to Automatic Melanoma Diagnosis," IEEE Access, vol. 8, pp. 31254–31269, 2020, doi: 10.1109/ACCESS.2020.2973188.

- [16] S. Jiang, H. Li, and Z. Jin, "A Visually Interpretable Deep Learning Framework for Histopathological Image-Based Skin Cancer Diagnosis," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 5, pp. 1483–1494, May 2021, doi: 10.1109/JBHI.2021.3052044.
- [17] M. Shorfuzzaman, "An explainable stacked ensemble of deep learning models for improved melanoma skin cancer detection," *Multimed. Syst.*, vol. 28, no. 4, pp. 1309–1323, Aug. 2022, doi: 10.1007/s00530-021-00787-5.
- [18] L. Wei, K. Ding, and H. Hu, "Automatic Skin Cancer Detection in Dermoscopy Images Based on Ensemble Lightweight Deep Learning Network," *IEEE Access*, vol. 8, pp. 99633–99647, 2020, doi: 10.1109/ACCESS.2020.2997710.
- [19] X. Wang, X. Jiang, H. Ding, Y. Zhao, and J. Liu, "Knowledge-aware deep framework for collaborative skin lesion segmentation and melanoma recognition," *Pattern Recognit.*, vol. 120, p. 108075, Dec. 2021, doi: 10.1016/j.patcog.2021.108075.
- [20] A. A. Adegun and S. Viriri, "Deep Learning-Based System for Automatic Melanoma Detection," *IEEE Access*, vol. 8, pp. 7160–7172, 2020, doi: 10.1109/ACCESS.2019.2962812.
- [21] M. Manzo and S. Pellino, "Bucket of Deep Transfer Learning Features and Classification Models for Melanoma Detection," *J. Imaging*, vol. 6, no. 12, p. 129, Nov. 2020, doi: 10.3390/jimaging6120129.
- [22] S. İlkin, T. H. Gençtürk, F. Kaya Gülağız, H. Özcan, M. A. Altuncu, and S. Şahin, "hybSVM: Bacterial colony optimization algorithm based SVM for malignant melanoma detection," *Eng. Sci. Technol. Int. J.*, vol. 24, no. 5, pp. 1059–1071, Oct. 2021, doi: 10.1016/j.jestch.2021.02.002.
- [23] R. V. Selvarani and P. S. H. Jose, "A Label-Free Marker Based Breast Cancer Detection using Hybrid Deep Learning Models and Raman Spectroscopy," *Trends Sci.*, vol. 20, no. 4, p. 6299, Jan. 2023, doi: 10.48048/tis.2023.6299.
- [24] W. Gouda, N. U. Sama, G. Al-Waakid, M. Humayun, and N. Z. Jhanjhi, "Detection of Skin Cancer Based on Skin Lesion Images Using Deep Learning," *Healthcare*, vol. 10, no. 7, p. 1183, Jun. 2022, doi: 10.3390/healthcare10071183.
- [25] R. Fan, X. Li, S. Lee, T. Li, and H. L. Zhang, "Smart Image Enhancement Using CLAHE Based on an F-Shift Transformation during Decompression," *Electronics*, vol. 9, no. 9, p. 1374, Aug. 2020, doi: 10.3390/electronics9091374.
- [26] M. Dhanushree, R. Priyadharsini, and T. Sree Sharmila, "Acoustic image denoising using various spatial filtering techniques," *Int. J. Inf. Technol.*, vol. 11, no. 4, pp. 659–665, Dec. 2019, doi: 10.1007/s41870-018-0272-3.
- [27] L. Hu, Y. Zhang, K. Chen, and S. Mobayen, "A COMPUTER-AIDED melanoma detection using deep learning and an improved African vulture optimization algorithm," *Int. J. Imaging Syst. Technol.*, vol. 32, no. 6, pp. 2002–2016, Nov. 2022, doi: 10.1002/ima.22738.
- [28] Q. Abbas, F. Ramzan, and M. U. Ghani, "Acral melanoma detection using dermoscopic images and convolutional neural networks," *Vis. Comput. Ind. Biomed. Art*, vol. 4, no. 1, p. 25, Dec. 2021, doi: 10.1186/s42492-021-00091-z.
- [29] H. Nahata and S. P. Singh, "Deep Learning Solutions for Skin Cancer Detection and Diagnosis," in *Machine Learning with Health Care Perspective*, V. Jain and J. M. Chatterjee, Eds., in *Learning and Analytics in Intelligent Systems*, vol. 13. Cham: Springer International Publishing, 2020, pp. 159–182. doi: 10.1007/978-3-030-40850-3\_8.
- [30] N. Nigar, M. Umar, M. K. Shahzad, S. Islam, and D. Abalo, "A Deep Learning Approach Based on Explainable Artificial Intelligence for Skin Lesion Classification," *IEEE Access*, vol. 10, pp. 113715–113725, 2022, doi: 10.1109/ACCESS.2022.3217217.

# Enhancing Diabetic Retinopathy Detection Through Machine Learning with Restricted Boltzmann Machines

Dr. Venkateswara Rao Naramala<sup>1</sup>, B.Anjanees Kumar<sup>2</sup>, Dr. Vuda Sreenivasa Rao<sup>3</sup>, Dr. Annapurna Mishra<sup>4</sup>,  
Shaikh Abdul Hannan<sup>5</sup>, Prof. Ts. Dr. Yousef A. Baker El-Ebiary<sup>6</sup>, R. Manikandan<sup>7</sup>

Professor, CSE (AI&ML), RVR & JC College of Engineering, Chowdavaram, Guntur, Andhra Pradesh<sup>1</sup>  
Associate Professor, Electronics and Communication Engineering, Dadi Institute of Engineering and Technology,  
Anakapalle, Andhra Pradesh, India<sup>2</sup>

Associate Professor, Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India<sup>3</sup>  
Associate Professor, Electronics and Communication Engineering, Silicon Institute of Technology,  
Bhubaneswar, India, Pincode: 7510244

Assistant Professor, Department of Computer Science and Information Technology, Albaha University,  
Albaha, Kingdom of Saudi Arabia<sup>5</sup>

Professor, Faculty of Informatics and Computing, UniSZA University, Malaysia<sup>6</sup>.  
Research Scholar, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology,  
Avadi, Chennai, Tamil Nadu, India-600062<sup>7</sup>

**Abstract**—Diabetes is a potentially sight-threatening condition that can lead to blindness if left undetected. Timely diagnosis of diabetic retinopathy, a persistent eye ailment, is critical to prevent irreversible vision loss. However, the traditional method of diagnosing diabetic retinopathy through retinal testing by ophthalmologists is labor-intensive and time-consuming. Additionally, early identification of glaucoma, indicated by the Cup-to-Disc Ratio (CDR), is vital to prevent vision impairment, yet its subtle initial symptoms make timely detection challenging. This research addresses these diagnostic challenges by leveraging machine learning and deep learning techniques. In particular, the study introduces the application of Restricted Boltzmann Machines (RBM) to the domain. By extracting and analyzing multiple features from retinal images, the proposed model aims to accurately categorize anomalies and automate the diagnostic process. The investigation further advances with the utilization of a U-network model for optic segmentation and employs the Squirrel Search Algorithm (SSA) to fine-tune RBM hyperparameters for optimal performance. The experimental evaluation conducted on the RIM-ONE DL dataset demonstrates the efficacy of the proposed methodology. A comprehensive comparison of results against previous prediction models is carried out, assessing accuracy, cross-validation, and Receiver Operating Characteristic (ROC) metrics. Remarkably, the proposed model achieves an accuracy value of 99.2% on the RIM-ONE DL dataset. By bridging the gap between automated diagnosis and ophthalmological practice, this research contributes significantly to the medical field. The model's robust performance and superior accuracy offer a promising avenue to support healthcare professionals in enhancing their decision-making processes, ultimately improving the quality of care for patients with retinal anomalies.

**Keywords**—Optic disc (OD); Optic cup (OC); U-network; restricted Boltzmann machines; squirrel search algorithm

## I. INTRODUCTION

Diabetes mellitus, a common metabolic problem characterized by increased blood sugar levels, can cause diabetic retinopathy, a chronic and progressive eye condition. The retina, the light-sensitive tissue at the back of the eye in charge of conveying visual information to the brain, is especially impacted by this disorder that poses a threat to eyesight [1]. The primary cause of adult blindness globally is diabetic retinopathy, which is of great public health concern. The slow destruction of the tiny blood vessels that supply the retina is the defining feature of diabetic retinopathy. Long-term exposure to high blood glucose levels, a characteristic of diabetes, can erode and potentially harm these vulnerable blood vessels. As a result, they can start to leak, causing fluid and lipid deposits to build up in the retina. Alternately, they may narrow and reduce the retina's blood supply, causing the release of growth hormones that promote the development of aberrant blood vessels. Because these aberrant blood vessels are weak and prone to bleeding, the retina might sustain additional harm.

In order to properly manage diabetic retinopathy, early identification and prompt care are essential. Regular eye exams and specialized imaging methods, such as fundus images and optical coherence tomography (OCT), are crucial for identifying this problem and tracking its development. The development of automated methods for diabetic retinopathy screening has also been aided by developments in artificial intelligence and machine learning, which have the potential to enhance the accessibility and effectiveness of early diagnosis and treatment. Understanding and resolving the complexity of diabetic retinopathy remain crucial in preventing vision loss and raising the quality of life for people with diabetes as diabetes incidence rises internationally [2]. Recently, automated diabetic retinopathy identification using retinal

scans has shown promise thanks to machine learning. Machine learning algorithms are capable of classifying the severity of diabetic retinopathy with amazing accuracy after analysing large datasets of retinal images and extracting pertinent information. Achieving consistently excellent performance across various patient demographics and image quality situations still presents a number of obstacles. The diagnostic accuracy of traditional machine learning algorithms can be hampered by the difficulty with which complicated, hierarchical patterns in retinal images can be captured. This has motivated academics to investigate more sophisticated deep learning techniques, such as CNNs and RBMs, to improve the identification of diabetic retinopathy.

The retina was a sphere located on the retina's inner side at the eye's rear. Its purpose has been to interpret visual data using the cones and rod and other image receptors found in the eye. The macula, or retina's central region, has a black, rounded area. Sharp eyesight was provided by the macula's fovea, which is located at its centre. The body's circulatory supplies blood to the retinal cells, just like it does for any other type of tissue. Each of the veins and main arteries enter and depart the eye through the optical disc, which is formed by an optic cup. It also serves as an organ wherein the optical nerve exits the eye. The efficacy of treatments in the medical sector is increased by early illness identification. Whenever the pancreas doesn't create enough insulin, diabetes looks to be a disorder that becomes progressively. A severe public health issue, diabetes mellitus currently affects 463 million individuals globally and is expected to reach 700 million in 2045. Diabetic retinopathy (DR), a commonly prevalent eye condition associated with diabetes, affects at least 1/3<sup>rd</sup> of people who have the condition [3]. No matter how severe their diabetes has been every patient with diabetes may acquire DR, which is characterised by increasing vascular disturbances in their retinas. Non-proliferative DR (NPDR) and PDR are the two common stages of DR [3]. Among the most popular clinical measures, the International Clinical DR (ICDR) dimension, has five DR severity levels: regular, mild, serious, moderate, and excessive [4]. There has been thought to be 93 million individuals living with DR globally, making it the main causes of blindness amongst working-age individuals. These numbers are expected to rise even higher, largely as a result of rising prevalence of diabetes in emerging Asian countries like China and India [4]. Whereas the early phases of DR are typically asymptomatic, neuronal retinal injury and clinically undetectable microvascular alterations advance during this time [5]. As a result, people with diabetes should undergo routine eye exams because prompt identification and treatment of the problem are crucial [4]. Flashes and floaters, vision loss, and blurred vision have been the main signs and manifestations of DR [4]. The macula is impacted by DR, which is brought on by metabolic changes inside retinal blood cells, caused because of abnormal blood flow, and blood components and blood leaks throughout the retina. As a result, the retinal membrane swells and causes vision to become hazy or blurry. The condition damages both eyes, and with continued untreated diabetes for an extended period of time, diabetic maculopathy results in blindness [6]. Early recognition of DR has been even more crucial given that the primary preventive approach was the treatment of

hyperlipidemia, hypertension, and hyperglycemia [5]. Additionally, if diabetic maculopathy and proliferative retinopathy have been cured in the initial phases of the illness, there is a considerable reduction in the risk of blindness with currently accessible therapies, including laser imageocoagulation. It becomes clear that early identification and suitable therapy are crucial for delaying or even preventing blindness from DR [7].

With the aid of specific medically recognised and authorised comparing agents that have been inserted towards the retina, the pupil elongation occurs in order to monitor the retinal structure, including the RBVs, macula, fovea, and Optic Disc (OD). This approach makes use of a mydriatic foundation video or the fluorescein catheterization (FA). This helps to collect fundus images from people with diabetes so they may be examined for the accurate diagnosis as well as the detection of DR. Manual techniques, including bio-microscopy, fundus retinal images, Optical Coherence Tomography (OCT), Adaptive Optics, Retinal Oximetry, Scanning Laser Ophthalmoscopy (SLO), OCT Angiography, Doppler OCT, Retinal Thickness Analyzer (RTA), and many others, can be used to identify DR early on [8]. Nevertheless, the manual analysis of the disease using these traditional approaches is laborious, time-consuming, and extremely error-prone. Additionally, it necessitates an advanced task force, which is occasionally impractical given the current situation. Therefore, it is not possible to carry out manual identification for DR early identification at any moment or location. As a result, the magnitude of these difficulties has spurred extensive study and the creation of methods that instinctively analyse data and detect the DR onset. This is mainly due to the desire to lower the expense of treating the illness. The common objective of increasing the effectiveness of healthcare services has been supported by developments in medical technology [9]. e-Health structures, for instance, have been effective being employed in an assortment of healthcare routes [10], [11]. In the biomedical imaging discipline, computer vision-oriented systems are becoming more significant. They offer the radiologist valuable decision support data that improves the prognosis and effectively informs medical personnel on the optimum efficient therapies for important medical diseases. Various image modalities, including 7-field colour fundus images (CFPs), ultra-wide-field-SLO (UWF-SLO), and colour fundus images (CFIs), were employed for the evaluation and therapy of DR in the particular medical imaging realm. Benefits could include less labour for the ophthalmologist as well as a lower risk of human miscalculation. Additionally, compared to manual detection, a computerised system might be far more effective and easily able to spot lesions and anomalies. Automated of DR identification is therefore crucial. Artificial Intelligence (AI) techniques can be used to create the DR automation systems.

Development in research and innovation has improved the quality, safety, and liveability of individual life. For example, a cutting-edge field of study called automatic medical imaging assessment uses imaging technologies to spot dangerous disorders. These Computer-Aided Diagnostic (CAD) tools, also known as automatic diagnosis systems (ADSSs), offer

amenities for the benefit of people. The use of CAD systems was acknowledged in a wide range of clinical diagnoses, including the identification of brain tumours, DR, gastrointestinal problems, glaucoma, macular oedema, and many others [12]. With personal observations, such CAD technologies have generated outcomes that are comparable. The creation of CAD tools depends on the image attributes that are computationally retrieved from images [13]. Clinical staff must manually examine DR abnormalities and retinal anatomy characteristics. Furthermore, observing intra- or inter-changes in retinal elements needs technical domain expertise because human examination of DR has been subjective. As a consequence, earlier DR screening programmes with colour fundus images must be carried out by CAD technologies. Clinical staff may spend less time, money, and effort manually analysing retinal images thanks to these CAD tools [14], [15]. AI offers the ability to completely alter how eye diseases have been now diagnosed and have a profound therapeutic effect on advancing ophthalmic treatment. Prior research on automatic DR identifications has been conducted. A useful method for the automatic identification and evaluation of DR has grown up: deep analysis of fundus images. The patients in physically underserved regions where there are few ophthalmologists and scarce medical facilities could benefit from the efficient deep learning (DL) system's ability to accurately and automatically recognise more severe DR with comparable or superior reliability than learned instructors and retina experts [16].

Machine Learning (ML) is a subgroup of DL that has grown stronger and increasingly effective as a weapon. DL is also an ideal complement to ML. The DL architecture is made up of a multifaceted, hierarchical design. Classifying, segmenting, localising, and identifying medical images are major DL tasks in clinical image assessment. DL offers more promising and outstanding outcomes in the diagnosis and categorization of DR diseases using a variety of techniques. Convolutional Neural Networks (CNN), Auto encoders, Generative Adversarial Networks (GAN), Recurrent Neural Networks (RNN), Deep Boltzmann Machines (DBM), Deep Belief Networks (DBN), and Deep Neural Networks (DNN) have been a few examples of DL-based approaches. The model performs better when there has been additional training information since both low-level and high-level characteristics have been automatically retrieved and trained [17]. Convolutional Neural Networks (CNN), a fundamental framework of DL in computer-vision, have produced outstanding results with regard to forecasting and detection in the medical images categorization. By autonomously segmenting the region-of-interests (ROIs) from the initial images, DL methods streamline the feature retrieval procedure. Additionally, these methods offer a complete answer for developing and assessing the categorization model. In DL methods, The DR images have been initially gathered. The acquired DR images have been then subjected to preprocessing methods like contrast enhancement, lighting modification, and scaling to remove noisy features. After being pre-processed, those images have been sent to the DL design, which uses the learning images' distinctive characteristics as input and the retrieved attributes and their scores to acquire categorization rules. In DL training, the

attribute weights have been recursively optimised to get the optimal characteristics weight and more precisely categorise the images. These optimised weights have been finally examined on unlabelled images via a categorization layer. The use of cloud computing has permeated almost every element of human life. Cloud computing, on the other hand, has limits in terms of prolonged delays, which are detrimental to IoT activities that require real-time responses [18]. The key contribution is discussed as follows:

- The research highlights the critical importance of early diagnosis for diabetes-related retinal conditions, which can potentially lead to blindness if not identified and managed in a timely manner.
- The paper recognizes the challenges posed by labour-intensive and time-consuming traditional methods of diagnosing diabetic retinopathy and the subtlety of early glaucoma symptoms. It underscores the need for more efficient diagnostic techniques.
- The study integrates machine learning and deep learning techniques to address these challenges. Notably, the introduction of Restricted Boltzmann Machines (RBM) showcases an innovative approach to enhance the accuracy and efficiency of the diagnostic process.
- The proposed model extracts and analyses multiple features from retinal images using RBMs, facilitating the accurate categorization of anomalies. This approach holds the potential to significantly expedite the diagnostic process.
- The research further extends its impact by introducing a U-network model for optic segmentation, contributing to more precise and reliable diagnostics.
- The application of the SSA to fine-tune RBM hyper parameters demonstrates a commitment to achieving optimal model performance.
- Rigorous experimentation using the RIM-ONE DL dataset substantiates the effectiveness of the proposed methodology. The high accuracy achieved serves as a testament to the model's robustness and potential clinical relevance.
- The comprehensive comparison against previous prediction models, incorporating metrics like accuracy, cross-validation, and ROC, underscores the model's superiority in diagnostic capabilities.

The structure of this article is organized as follows: Section II reviews previous research on prediction problems using various optimization methodologies. Section III, image pre-processing improve and adjust the brightness. The OD and OC are segmented using a Threshold U-Network. Section IV discusses the rim-one-dl dataset results. Section V, experimental evaluation comprises mathematically developed system models for sensitivity, specificity, accuracy, and AUC. Section VI concludes the paper.

## II. RELATED WORKS

A computer vision-based method to analyse and forecast diabetes from input retinal images was proposed by Mini Yadav et al. [19]. This facilitates the early identification of DR. Pre-processing, feature extraction, and segmentation techniques are used in the aforementioned image processing stage. Following the image processing procedures, the categorization step using ML is carried out. Python is employed for better research outcomes. For designing the coding for the empirical outcomes' platform, investigators use Jupyter. The framework created was tested on databases from open access public repositories, and it used CNN to obtain 98.50% accuracy as opposed to SVM's 87.40% accuracy. These outcomes outperform a number of sophisticated unsupervised ML approaches. It leads to a reduction in procedure complexity and enhanced evaluation metrics, making it appropriate for application in the DR detection employing retinal image assessment. The study asserts that their unsupervised ML approach outperforms previous methods, but in order to judge the robustness and generalizability of the suggested method, it must be tested on independent databases. The dependability of the outcomes might not be ensured in the absence of external testing.

The multi-scale attention network (MSA-Net) was proposed by Mohammad and Yasmine [20] for DR categorization. The suggested method uses an encoding network to place the retinal images in an upper-level representational space, enriching the representation by combining middle- and the highest-level information. The retinal morphology in a distinct locale is then described using a multi-scale characteristic hierarchy. A multi-scale attention system is added on the highest of the upper-level structure in order to improve the feature depiction's ability to discriminate. The cross-entropy loss has been employed to learn the framework in a conventional manner, and it is used to categorise the DR seriousness level. In addition, the model is instructed to distinguish between healthy and unhealthy retinal images employing the poorly annotated information. This stand-in exercise aids the model in improving its ability to distinguish between unhealthy retinal images. When used, the suggested strategy produced remarkable outcomes on the APTOS and EyePACS public databases. The study discusses the application of poorly labelled data to help the algorithm distinguish among retinal images that are healthy and those that are ill. However, if the annotations have not been precise or thorough adequate to capture the entire spectrum of abnormal retinal images, the efficacy of this method may be constrained. The training process may become biased and noisy due to insufficient annotations.

A CNN-oriented strategy was suggested by Worapan et al. [21] as a means of identifying and evaluating DR from retinal images. It could categorise the retinal images that were supplied into a normal group or atypical group that would then be automatically divided into four levels of anomalies. The suggested solution has been predicated on DeepRoot, a recently suggested CNN framework. It has one major branch and two subsidiary branches that link it. The principal feature generator for both upper- and lower-level attributes in retinal images resides in the central branch. The characteristics

produced from the central branch are then subsequently extracted by additional branches to provide more intricate and comprehensive characteristics. They use customised zoom-in/out and attenuation stratum to record the fine details of minuscule remnants of DR within retinal images. Moreover, the Kaggle database is used to train, verify and test the suggested approach. Using examples of unknown information that the researcher independently acquired from an actual circumstance in a hospital, the learned model's regularisation is assessed. Under the two categories scenario, it performs admirably, with 98.18% sensitivity. Although the research suggests evaluating the regularisation of the trained model using samples of unknowable information obtained from a healthcare facility, the specifics and extent of this verification have not been given. To guarantee that the model's performances comply with the desired purpose and to gauge its efficacy in practical contexts, rigorous real-world verification studies must be carried out in clinical settings.

Employing CNN, Raja and Balaji [22] created a reliable automated diabetic identification system for retinal images. The proposed system primarily consists of five components like preprocessing, exudates segmentation, blood vessels segmentation, extraction of textural features, and diagnosis of diabetes. The input retinal image is initially enhanced during the preprocessing stage employing adaptive histogram equalisation (AHE). As a result, fuzzy c-means clustering (FCM) and CNN have been employed for segmenting exudates and blood vessels, respectively. Following that, exudates and blood vessels are used to extract textural properties. Following extracting features, a support vector machine (SVM) is used to classify diabetic patients. The experimental findings show that, when compared to other methodologies, the suggested approach achieves improved diabetic diagnosis outcomes (sensitivity, specificity, and accuracy). To guarantee the suggested method's suitability for various populations and changes in retinal images, its efficacy and adaptability should be tested on separate databases. If various datasets were utilised to show the system's durability, the assessment should have been done on many databases, and this should be specified in the study.

The likelihood of vision loss may be considerably reduced by early diagnosis and treatment with DR. Analysis of retinal image has gained popularity as a method of disease detection in contemporary ophthalmology. Fundus angiography has been frequently used by ophthalmologists and computerised systems to find DR-based clinical indicators for DR earlier detection. Because of the imaging situations complexity, like imaging in a range of angles and lighting circumstances, fundus images are frequently exposed to inadequate contrast, sound, and inconsistent illumination problems. By lowering the noise and boosting the contrast, Saif Hameed et al. [23] proposed an algorithm for enhancing image quality that will reinforce the norm for colour fundus imaging. The method involves two primary steps: shrinking the images to eliminate extraneous information, followed by performing the shape cropping as well as gaussian smoothing to boost contrast and reduce noise. MESSIDOR and EyePACS represent two common datasets used to assess the research outcomes. It has been amply demonstrated that the findings of improved image

feature retrieval and categorization surpass those obtained without using the enhancement technique. As an IoMT usage, the updated algorithm has also been evaluated in smart hospitals. Although the research includes evaluating the revised method in smart hospitals, it is vital to take into account the practical difficulties and prerequisites for clinical deployment in the actual world. This entails integration with currently in use medical imaging equipment, regulatory authorizations, and professional acceptability. These issues should be covered in the study together with information on the difficulties and possible clinical advantages of applying the method in a clinical context.

Anas Bilal et al. [24] proposed an innovative two-stage method for automatic DR categorization. Data augmentation and pre-processing methods have been utilised to improve the image clarity and quantity because the asymmetric blood vessels (BV) and Optic Disc (OD) identification system has a low percentage of positive cases. For BV and OD segmentation during the first stage, two distinct U-Net algorithms are used. Subsequent to preprocessing for extracting and selecting the most discriminating characteristics after BV and OD extraction employing Inception-V3 depending on transfer learning (TL), the symmetrical integrated CNN-SVD framework has been developed in the secondary phase. It recognises DR by identifying retinal indicators like exudates, haemorrhages, and microaneurysms. The suggested approach displayed cutting-edge efficiency on the Messidor-2, DIARETDB0, and EyePACS-1 tests, with mean accuracy readings of 94.59 per cent, 93.52 per cent, and 97.92 per cent, correspondingly. The effectiveness of the recommended approach has been demonstrated through extensive evaluations and comparability with benchmark methods. The low affirmative case prevalence for asymmetric BV and OD recognition can make it difficult to train and assess the algorithms. The accuracy and generalizability of the suggested strategy, especially for identifying certain abnormalities connected to DR, may be impacted by the dearth of positive instances.

Phridviraj et al. [25] presented a bi-directional LSTM-oriented DR diagnosis model employing retina fundus images. The Bi-directional LSTM approach uses retinal fundus images to identify and categorise various categories of DR. The Multiscale Retinex alongside Chromaticity Preservation (MSRCP) technique is used in the suggested framework as a preprocessing stage to raise fundus image disparity and advance short discrepancy of pharmaceutical perspectives. Moreover, MSRCP is utilised to create results that are adequate for processing images. The key challenge regarding the Retinex method is setting the values for the variables, including the gain, offset, Gaussian scales, etc. Such parameters need to be altered to produce a useful result. The major objective of the presented approach is to get the ideal settings for the MSRCP procedure's variables. Additionally, an effective net-based characteristic extraction that makes of DL is used to create feature vectors from already processed images. The standard MESSIDOR database is used in numerous experiments. The outcomes are examined in light of several evaluating criteria. The findings demonstrate that the Bi-LSTM-MSRCP methodology is superior to more

contemporary techniques for DR diagnosis. Even if the research shows better results than current methods, it's crucial to take the comprehension of the suggested model into account. Since bi-directional LSTM models are frequently regarded as "black-box" models, it might be difficult to comprehend how decisions are made and the particular characteristics that contribute to DR diagnosis. The clinical acceptance and confidence in the paradigm may be constrained by its lack of comprehension.

For the purpose of DR identification, Loheswaran [26] suggested the Optimised Kernel-oriented Fuzzy C-Means (OKFCM) based Segmentation and RNN-based Categorization scheme. OD elimination and KFCM Separation performed using Modified Ant Colony Optimisation (ACO) are the two primary stages within the recommended segmentation portion. GLCM and time-built characteristics have been utilised for the grading of DR. The OKFCM-MACO-RNN was another name for the suggested framework. The OKFCM-MACO-RNN technique's sensitivity, accuracy, and specificity have values of 81.65 percent, 99.33 percent, and 99.42 percent, accordingly, have been adjusted to finish the OKFCM-MACO-RNN categorization assessment procedure in the diaretDB1 database. The OKFCM-MACO-RNN approach can be relied upon to reliably detect exudates. The OKFCM-MACO-RNN based Segmentation are evaluated with jacquard coefficient, accuracy, and dice coefficient having values of 72.84, 85.65, and 93.15 correspondingly. The quantity and variety of the database employed for learning and evaluation have a significant impact on how well the suggested framework performs. The diaretDB1 dataset has been mentioned in the study, but it is crucial to evaluate the dataset's representativeness and determine whether it effectively accounts for the variety of DR situations that are experienced in real-world circumstances. The algorithm might not work as well on other databases.

It is quite challenging to diagnose DR through laborious physical diagnosis because of the variety and complexities of DR. As a result, Spoorthi and Rekha [27] concentrated on employing an RNN-LSTM and Deep CNN (DCNN) integration to categorise a specific collection of fundus images into four phases. This method recognises all DR phases instantly. This mixture takes a lot of details from the fundus image. To create the integrated model, over 2000 fundus images have been utilised overall. This work shows that the precision of forecasting the DR phases has been greatly increased by the retrieval of those characteristics from fundus images employing RNN-LSTM and DCNN. To determine the combined model's suitability for use with various populations and changes in retinal images, its efficacy and generalizability should be examined on separate databases. To show the approach's resilience and dependability across various datasets, additional verification is required.

### III. PROBLEM STATEMENT

This study focuses on improving diabetic retinopathy diagnostic accuracy and efficiency using machine learning and RBMs. Diabetic retinopathy is a major consequence of diabetes that, if not identified early, can result in blindness.

Current approaches for detecting diabetic retinopathy frequently rely on ophthalmologists doing manual assessments, which can be expensive, time-consuming, and prone to inter-observer variability. Additionally, existing automated methods sometimes have trouble recognizing retinal abnormalities with high sensitivity and specificity. The suggested strategy tries to use RBMs, a class of neural network capable of learning hierarchical representations of data, to improve the automated identification of diabetic retinopathy in order to overcome these constraints. By doing this, it hopes to provide an alternative to current techniques that is more precise, effective, and consistent [24].

However, the limits of existing diabetic retinopathy detection techniques call for improvements. First off, the diagnostic accuracy of traditional machine learning algorithms is sometimes constrained by their inability to grasp the complex patterns and subtle elements seen in retinal images. Additionally, the lack of extensive, diverse datasets for training and testing might limit the capacity of present models to generalize, particularly when faced with differences in retinal images that occur in the real world. The promise of deep learning methods like RBMs, which can build intricate, hierarchical representations from data, is also underutilized in many current approaches. The need to build a strong, RBM-based solution that can not only exceed existing approaches in terms of accuracy but also adapt successfully to changes in image quality, illness development, and patient demographics is therefore appealing.

#### IV. PROPOSED METHODOLOGIES

Retinal image analysis is essential to perform mass screening, even if there is a lack of experienced doctors. This work presents RBM images, the analysis uses numerous features to predict formation in images. The steps used are image acquisition, pre-processing, segmentation, and glaucoma classification. Initially, retinal fundus images are obtained from the public data source. Pre-processing is necessary to eliminate unwanted image features such as speckles, blind spots etc. The OD and OC is performed using a Threshold U-Network (Two-stage TUNet). Next, the deep features such as contrast, homogeneity, mean disc, correlation, SD (standard deviation) disc, energy disc, entropy disc,

entropy cup, SD cup, mean cup, energy cup, etc., are extracted. Fig. 1 shows flow diagram of deep learning based retinal image analysis.

##### A. Dataset Collection

Some confusion and incorrect usage of the three published versions prompted the proposal to update and integrate DL (rim one for deep learning), an updated version of them, was created using them. With profound learning in mind, it is designed to meet the standards mentioned earlier for the environment. The outcome of integrating the three previous versions is rim-one DL. This gives each patient and each eye a separate image. Additionally, all images were correctly cropped around the head of the optic nerve, utilizing the exact same parallelism criterion, which was not done in earlier versions.

##### B. Pre-Processing

Pre-processing techniques such as cropping, channel separation, and data enhancement are used to input images. A image is the removal of external parts by cropping to improve encircling, change the proportions of an angle, or emphasize a subject and resizing image results in a change in its physical size. A tile is divided into the portion of data that is lost due to cropping. The original dataset contained only a limited number of images that could be used as training data. A neural network could not be applied and trained with such an amount of data. Flipping and data rotation has been used as augmentation techniques to overcome this problem. To increase the dimension of the test set, the images first had to be vertically flipped. Images are then flipped vertically, an additional increasing amount of data available. Flipped images were rotated from 0° to 180° in 20° increments after flipping. As a result, the model was effectively trained with a larger data set.

##### C. Segmentation of Images

The technique of splitting an image into its component sections using a segment/part of an image is known as image segmentation. The task of objects is to find a certain subdivision among the provided images. Segmentation is also used in retinal fundus images to investigate structures like the optic disc and optic cup.

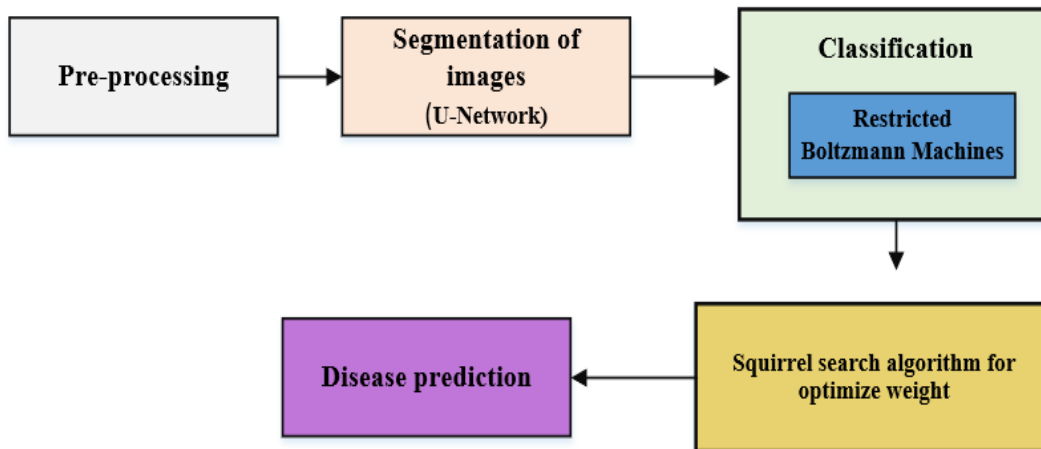


Fig. 1. Flow diagram of deep learning based retinal image analysis.



1) *OD*: The optic nerve head OD is where ganglion cells and axons leave eye to form visual cortex. If there are no shafts or cones, there is a blind zone. Because receptor rods and cones in this region are dull, the corresponding point of tiny blind spot at OD is increased. The OD is widely employed in to determine the cup-to-disc ratio.

2) *OC*: It is located in middle of the optic disc, which appears to be shaped like a white cup. The optical curve is responsible for the mobility of the retina in the human eye. The greater the distance from the optic curve's diameter, the greater the chance of developing glaucoma.

3) *Threshold U-network*: The OD and OC are segmented using a threshold U-network (TUNet). During processing, both OD and OC segmentation are two-stage thresholds [28]. A fundamental image processing approach is threshold segmentation. Varying tissues in medical imaging are often displayed at different levels of intensity. The threshold eliminates the uninteresting areas while also reducing noise interference. This process is both required and efficient. The threshold value must be chosen carefully. The region of interest will be impacted if the threshold range is too tiny; as a result, the experimental results are inaccurate. Noise reduction becomes difficult if the threshold range is set too high. The threshold range is often defined by experience. The two-stage U-Net design is used for the threshold interval computations. As glaucoma detection must be divided into two groups, it is challenging to segregate glaucoma data in real time (OD and

OC).To simplify and reduce the complexity of the problem, a two-stage U-Net framework is used and composed of total glaucoma segmentation and glaucoma substructure segmentation (the second step). The primary goal of the U-net architecture is the segmentation of the many substructures of the system [29]. Fig. 2 illustrates the U-Net framework.

*D. Restricted Boltzmann Machines (RBM) Diabetic Retinopathy Classification*

Using input vector  $f$ , concealed units  $s$ , weights matrices  $C$ , visible biases  $y$ , plus hidden biases  $z$  as shown, the RBM may be shown as a graph that is undirected.

$$L(s = 1|f) = \sigma(z + C.f)$$

$$L(f = 1|s) = \sigma(y + C.s) \tag{1}$$

The binary data acquired by applying the CLBP descriptors is transformed into actual value matrices in our process using RBMs. This information comprises image characteristics which the first set of CNN filters that have been trained under supervision often recognise, such as edges, blobs, or basic forms. Utilising RBMs enables bypassing these layers, leading to a reduced architecture to address a particular image categorization job. As a result, we are able to analyse more complicated features in the initial convolutional layers thanks to our prior processing and the RBM layers. Fig. 4 shows the restricted Boltzmann machine.

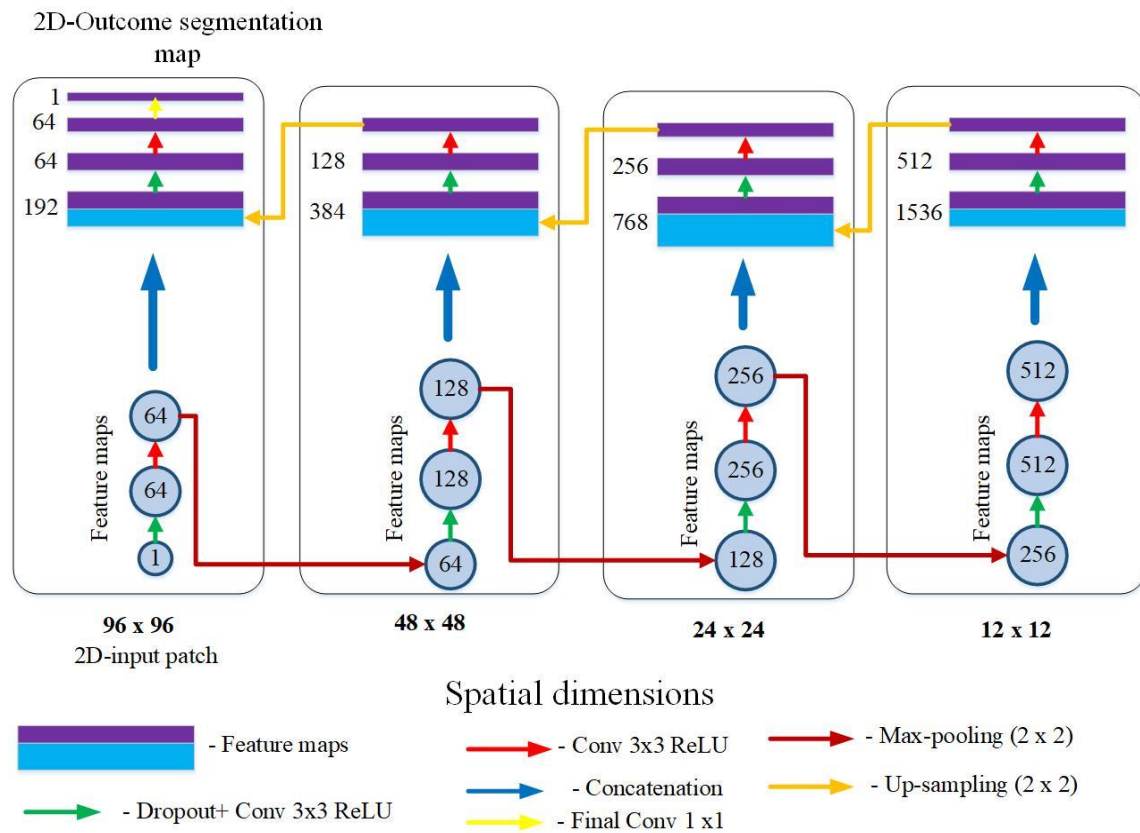


Fig. 2. U-net architecture.

The RBM analyses the information once the CLBP step is complete. The total amount of observable subunits is 16 in the most basic version since descriptor values are supplied immediately to the RBM, and bigger receptivity areas are possible by combining CLBP characteristics from a kernel of size P. The dimension of an output matrix as well as the overlapping sections may be modified using the stride length G. The RBM output vector  $v$  has a length of  $m = P^2$ .

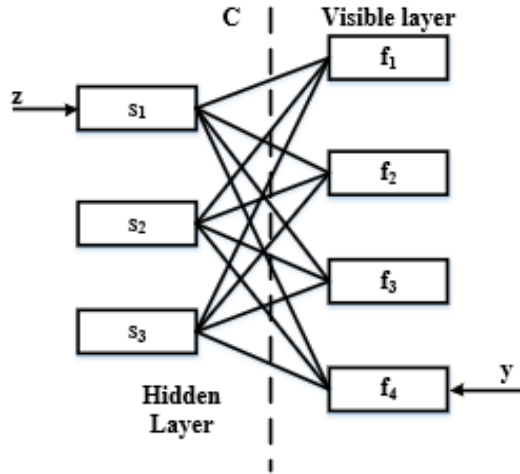


Fig. 3. Illustration of a restricted Boltzmann machine.

The input vector for the RBM is shown in Fig. 3 being created for  $P = 2$ . The RBM layer's feed is a matrix created as following.

$$RBM_{1/P} = \begin{bmatrix} f(1,1) & f(1,2) & \dots & f(1,m) \\ f(2,1) & f(2,2) & \dots & f(2,m) \\ \vdots & \vdots & \vdots & \vdots \\ f(m,1) & f(m,2) & \dots & f(m,m) \end{bmatrix} \quad (2)$$

The numerical value within the  $s_i$  component corresponding to the accessible vector  $f$  is given by because the RBM in this design de facto transforms into a typical neural network feeding forward with the activation function of sigmoid.

$$s_i = \sigma(z_i + \sum_j c_{ji} f_j) \quad (3)$$

In matrices recording, the final solution that describes the RBM level is

$$\sigma(Z + C \cdot RBM_{1/P}) \quad (4)$$

The scope that the vector produced by the preparation workflow relies on the number of hidden elements (RH) or cadence (S), which determines how multidimensional the resulting data is,

$$\left[ \frac{C}{G} \times \frac{S}{G} \times RH \right] \quad (5)$$

wherein, the supplied image's parameters are  $[C \times S]$ . The whole pretreatment workflow and the CLBP-RBM translation are shown in Fig. 4.

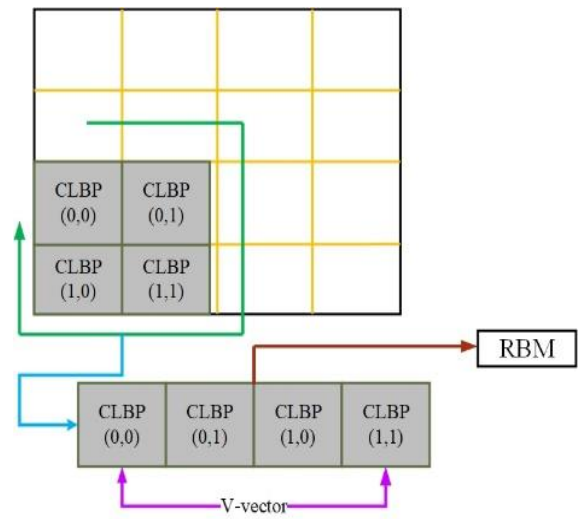


Fig. 4. CLBP kernels.

1) *Squirrel search algorithm for weight optimization*: The proposed method has 2 search methods: the first is jumping method, and second is a broadminded method. As part of evolutionary process, the practical method is automatically selected through the selection strategy of linear regression, which increases the robustness of a squirrel search algorithm. The escape process sufficiently creates a searching area, and the demise procedure uses the leaping mechanism in order to examine the created space. The capacity of SSA to evolve and explore is balanced. Regarding the incremental search approach, mutations operation maintains the present historical data and gives greater consideration to maintaining population diversity.

a) *Random initialization*: Initial population  $W$  was generated randomly with  $k$  number of flying squirrels.

$$W = \begin{bmatrix} W_{1,1} & W_{1,2} & \dots & \dots & W_{1,d} \\ W_{2,1} & W_{2,2} & \dots & \dots & W_{2,d} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ W_{m,1} & W_{m,2} & \dots & \dots & W_{m,d} \end{bmatrix} \quad (6)$$

In this case,  $W_{i,j}$  indicates the  $j^{th}$  dimension of  $i^{th}$  flying squirrel. In a forest, flying squirrels are allocated their initial locations according to a uniform distribution.

$$W_i = W_L + U(0,1) \times (W_U - W_L) \quad (7)$$

A distributed uniformly random number in the range  $[0, 1]$  is  $U(0,1)$ , and  $W_L$  and  $W_U$  represent lower and upper bounds for  $j^{th}$  dimension  $i^{th}$  flying squirrel.

b) *Fitness evaluation*: To calculate the localization, fitness values of the defined by users fitness coefficient has choice parameters (solution vectors) placed into it. The following table contains the values that were obtained:

$$f = \begin{bmatrix} f(|W_{1,1}, W_{1,2}, \dots, W_{1,d}|) \\ f(|W_{2,1}, W_{2,2}, \dots, W_{2,d}|) \\ \vdots \\ f(|W_{n,1}, W_{n,2}, \dots, W_{n,d}|) \end{bmatrix} \quad (8)$$

The value of fitness neural network layer weight depicts the quality of optimal weight. Here  $W = (w_t^c, w_t^{cap}, w_t^{GRU})$ , after storing the fitness values, array is sorted in ascending order. The squirrel with a minimum chestnut nut tree declares itself to have a fitness value. Afterwards, best three flying squirrels are thought to move to hickory nut trees from acorn trees.

c) *Generating location:*

Case 1: Hovering squirrels may move from acorn nut trees  $(W_{i,j})_{ant}$  to hickory nut trees  $(W_{i,j})_{hnt}$ . It is possible to determine the new location of squirrels in this case by following these steps:

$$(W_{i,j})_{ant} = \begin{cases} (W_{i,j})_{ant} + x_g \times s_c \times ((W_{i,j})_{hnt} - (W_{i,j})_{ant}) & \text{if } R_1 \geq P \\ \text{random location} & \text{otherwise} \end{cases} \quad (9)$$

Where  $x_g$  indicates a distance of random gliding,  $R_1$  indicates a random number in range of [0,1],  $(W_{i,j})_{hnt}$  denotes location of squirrel reached hickory nut tree and  $t$  represent current iteration. In mathematical model, the gliding constant  $s_c$  is used with the purpose of striking a balance amid exploration and extraction.

Case 2: A squirrel on a normal tree  $(W_{i,j})_{nort}$  can move towards an acorn tree; new location of the squirrel can be determined as follows:

$$(W_{i,j})_{nort} = \begin{cases} (W_{i,j})_{nort} + x_g \times s_c \times ((W_{i,j})_{ant} - (W_{i,j})_{nort}) & \text{if } R_2 \geq P \\ \text{random location} & \text{otherwise} \end{cases} \quad (10)$$

Case 3: In some cases, squirrels that have already consumed acorns on normal trees may settle in hickory trees to stockpile beech nuts which might be eaten during famine. Thus, the following information may be used to find new squirrel locations:

$$(W_{i,j})_{nort} = \begin{cases} (W_{i,j})_{nort} + x_g \times s_c \times ((W_{i,j})_{hnt} - (W_{i,j})_{nort}) & \text{if } R_3 \geq P \\ \text{random location} & \text{otherwise} \end{cases} \quad (11)$$

In this case,  $R_3$  represents a random number in the range of [0, 1]. The optimization algorithm was designed using an approximated model of the gliding behaviour. Flying squirrel fitness values are computed, and the locations are updated on

each iteration until the maximum number of iterations is reached.

V. RESULT AND DISCUSSION

The experimental setup, performance measurement, evaluation datasets, and experimental outcomes are all described in this section. The discussion of the results includes an evaluation of the proposed SSA algorithm. OD and OC were trained with the original data and merged with the image to demonstrate the effectiveness of the system based on deep feature extraction and classification.

A. *Simulation Environment*

Evaluate the proposed early-stage glaucoma disease prediction technique using the Python software in a simulated environment utilizing rim-one-dl datasets. The test is run on a machine equipped with an Intel(R) Core (TM) i5 CPU @ 3200 Mhz, 3 core(s), 4 logical pro. And micro software 10 pro, a micro soft corporation, is an OS manufacturer installed physical memory (RAM) 8GM. Table I shows simulation parameters for the rim-one-dl datasets.

TABLE I. SIMULATION PARAMETERS OF BOTH DATASETS

Simulation parameters for Rim-one-dl	Values
Batch size	2
Epoch	100
Activation function	Sigmoid
Dropout	0.5
Kernel size	(3,3)

B. *Performance Evaluation*

Fig. 5 compares the accuracy plot of the proposed ensemble DeepNet with conventional pre-trained models such as CapsNet, VGG 16 and DenseNet for the rim-one-dl datasets. An epoch in artificial neural networks is one loop that covers the whole training dataset. Training a neural network typically takes many epochs. The plot of CapsNet, shows a larger difference between validation and training accuracy. This shows that the model may be overfitted. Furthermore, the improvement in accuracy has not been constant. There are substantial gaps between training and validation at a few epochs. Furthermore, the training accuracy is significantly lower at the last epoch than the validation accuracy.

Furthermore, in the DenseNet plots in figures, the training accuracy is higher than the validation accuracy. This indicates that the model is overfitting, and the peaks are not increasing consistently. The training accuracy of DenseNet is continually rising; however, the validation accuracy is not. There is also a decrease in validation accuracy after a few epochs, which is undesirable. As a result, the model underperforms and is unsuitable for predicting abnormality.

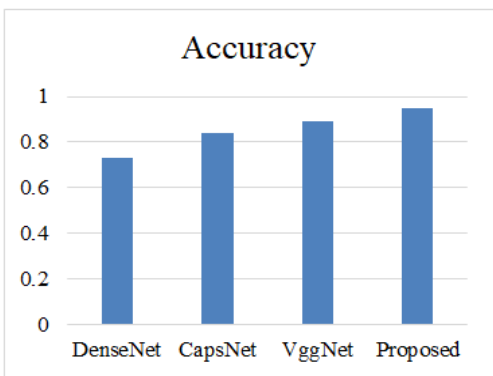


Fig. 5. (a) Accuracy plot.

Fig. 6 depicts the overall performance of several criteria, such as accuracy, specificity, recall, and precision, using the rim-one-dl datasets. It is based on deepnet ensembles such as deep CNN with multi-pooling layer, dual part scaled cross dense network, and graph CapsNet. Accurate is the ratio of correct predictions to the total prediction or true value. It is usually multiplied by 100 to express it as a percentage. It quantifies the accuracy of the classifier model and is scientifically calculated as

$$accuracy = \frac{TP + TN}{P + N} = \frac{TP + TN}{TP + TN + FP + FN} \quad (12)$$

Where  $TP$  denotes True positive,  $TN$  denotes True negative,  $FP$  denotes false positive and  $FN$  denotes false negative. Fig. 6 clearly shows that the proposed one provides the highest accuracy, which is determined to be the rim-one-dl dataset. In contrast, other classifiers provide significantly lower accuracy. In terms of specificity, it again takes the lead regarding values compared to other classifiers. In the KGD dataset, the Capsnet is found to provide an accuracy that is somewhat close to the designed accuracy. CapsNet is an ANN machine learning system that may be used to better simulate hierarchical connections. CapsNet’s slow training and testing, lacks of spatial information is its drawback. As a result, its performance is reduced compared to the proposed model.

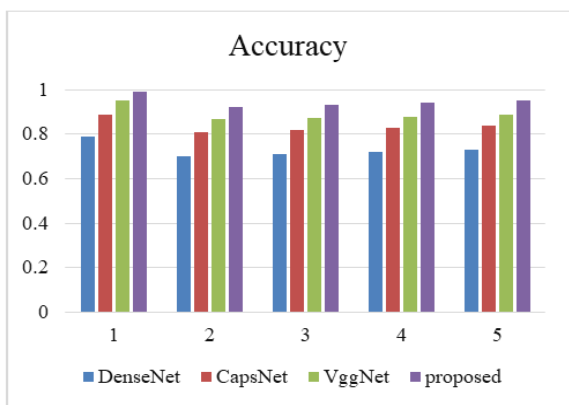


Fig. 6. Comparison graph for accuracy.

Also, a DenseNet is a convolutional neural network that uses dense connections between layers through Dense Blocks, which connect all levels directly. Each layer’s DenseNet feature maps are spliced with the previous layer, and the data is copied numerous times. However, the existing dense blocks did not consider the hierarchical features or concatenate the hierarchical features merely without removing the redundant features. As a result, its performance is not high as possible. The VGG16 network is large, implying it takes longer to train its parameters. Also, using the Max pooling layer in VGG16 caused information loss. The proposed model has tackled these issues using multi pooling layer, dual-part scaled cross-dense block and graph CapsNet. Hence, the proposed ensemble model outperforms all the pre-trained models shows in Fig. 6.

The rim-one-dl dataset utilized in the proposed research is one example of a dataset with varied comparison results that might vary due to a number of circumstances. The intrinsic diversity of data properties among datasets, including changes in image quality, resolution, noise levels, and the occurrence of certain retinal disorders, is a significant determining factor. Additionally, the design and training methods of the suggested algorithms might affect how well they function, which can make them more effective at managing particular kinds of data. On datasets with related features, algorithms designed for particular data qualities, such as those optimized for high-resolution retinal scans or skilled at managing noisy datasets, may perform better. Conversely, when applied to datasets with diverse properties, algorithms that lack the capacity to adapt to individual data changes may function inconsistently or with lower efficacy. The accuracy and generalizability of diabetic retinopathy detection systems can be enhanced by better understanding the specifics of each dataset and optimizing algorithms in accordance with those details.

The Jaccard Index is a statistical metric used to compare the similarity of samples. It is commonly referred to as the “Crossroads above the Union”. The Jaccard Index is the ratio of the intersection’s size to the samples’ union. It can have a minimum value of “0” and a maximum value of “1”. It is described mathematically as follows:

$$jaccard\ index, J(Gr, Sg) = \frac{Gr \cap Sg}{Gr \cup Sg} \quad (13)$$



Fig. 7. Segmentation performance.

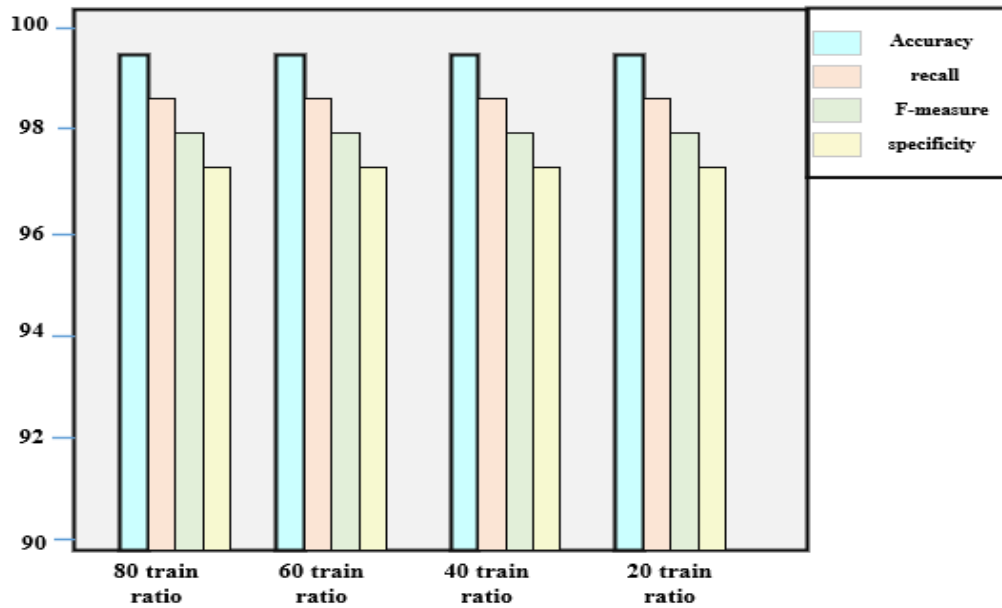


Fig. 8. Cross validation graph.

The k-fold cross-validation approach has several variants. The following are regularly used variants: Training and testing split: In the extreme, k can be set to 2 (rather than 1), resulting in a single train/test split for model evaluation. The cross validation for the rim-one-dl dataset and the KGD dataset is shown in the above Fig. 7. A typical procedure for testing the performance of a k-fold cross-validation model is the entire data set is randomly divided into independent k-folds without replacement. Fig. 8 shows the cross validation graph.

ROC curves in Fig. 9 were displayed to visually evaluate ensemble approaches based on the association between true positive and false positive rates for each deepNet. The graph shows better outcomes when such a curve is closer to the top-left corner. Using the rim-one-dl, the AUC value outperforms the existing Deepnet CapsNet, DenseNet, and VGG16. As a baseline, a random classifier is expected to give points along the diagonal (FPR = TPR). Below Fig. 9 shows the ROC curves of ensemble methods using deep CNN with multi pooling layer, dual part scaled cross dense network, and graph CapsNet.

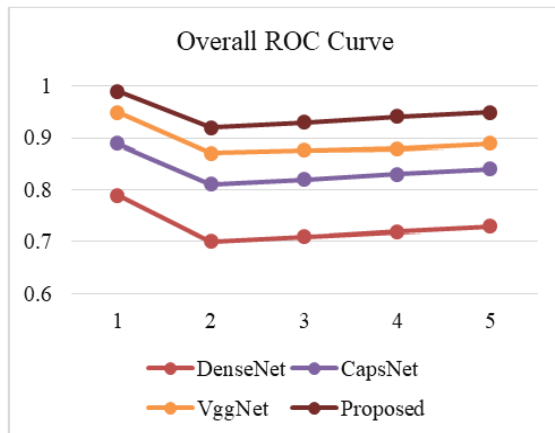


Fig. 9. ROC graph.

Plateaus or occasional dips might indicate challenges or local optima. Overall, the graph provides insight into the optimization's progress and efficiency in enhancing the solution's fitness over successive iterations.

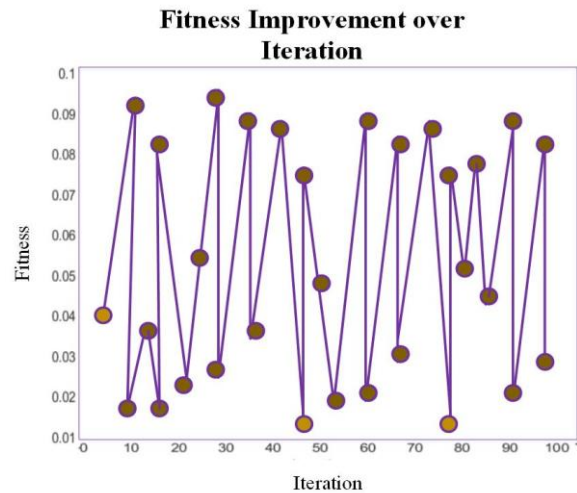


Fig. 10. Fitness improvement graph.

Above Fig. 10 illustrates how a certain fitness metric evolves across multiple iterations of an optimization process. The x-axis represents the iterations or time steps, while the y-axis represents the fitness metric. The graph demonstrates how the fitness value changes as the optimization algorithm refines its solution. Generally, the graph shows an initial steep improvement as the algorithm quickly converges towards better solutions, followed by a gradual slope as further improvements become harder to achieve.

C. Discussion

When compared to current techniques, the suggested method shows a lot of potential for the identification of

diabetic retinopathy. Traditional methods, which frequently rely on ophthalmologists' manual evaluation or crude automated algorithms, can be time-consuming, prone to human error, and unscalable in light of the expanding prevalence of diabetes worldwide. Contrarily, using the strength of RBMs is a state-of-the-art innovation that enables the automatic extraction of complex, hierarchical information from retinal images. This deep learning method offers a robust solution that can adapt to a variety of patient demographics and varied image quality situations, with the potential to significantly improve the accuracy and efficacy of diagnosing diabetic retinopathy. This suggested method, which uses RBMs, not only performs better than current methods but also represents a substantial advance in the early and accurate diagnosis of diabetic retinopathy, thereby assisting in the preservation of eyesight and the efficient use of healthcare resources.

## VI. CONCLUSION

In conclusion, this research presents a promising approach for automated diagnosis of diabetic retinopathy using restricted Boltzmann machines (RBMs) and retinal fundus images (RFIs). The proposed model incorporates various elements of prediction in retinal images, with a particular focus on the segmentation of using a thresholds U-network model. This segmentation step is crucial for identifying specific regions of interest in the retinal images and enables the subsequent classification process. To optimize the performance of the RBM, the research employs the Squirrel search algorithm (SSA) for selecting optimal hyperparameters and minimizing the weight of the RBM. The use of SSA helps improve the overall accuracy and efficiency of the automated diagnosis system. The evaluation of the model on the Rimone-dl dataset showcases promising results, achieving an accuracy of 99.2%. This high level of accuracy demonstrates the potential of the proposed system in accurately categorizing anomalies and aiding ophthalmologists in early and precise prediction of diabetic retinopathy. Future research may look at the integration of deep learning architectures, like CNNs, to increase feature extraction from retinal images and better classification accuracy for improving diabetic retinopathy diagnosis using machine learning using Restricted Boltzmann Machines. An interesting research direction may also involve investigating the possibility of federated learning and transfer learning techniques to use large-scale, heterogeneous datasets for better model generalisation in actual clinical scenarios.

## REFERENCES

- [1] U. Ishtiaq, S. Abdul Kareem, E. R. M. F. Abdullah, G. Mujtaba, R. Jahangir, and H. Y. Ghafoor, "Diabetic retinopathy detection through artificial intelligent techniques: a review and open issues," *Multimed. Tools Appl.*, vol. 79, no. 21–22, pp. 15209–15252, Jun. 2020, doi: 10.1007/s11042-018-7044-8.
- [2] W. L. Alyoubi, W. M. Shalash, and M. F. Abulkhair, "Diabetic retinopathy detection through deep learning techniques: A review," *Inform. Med. Unlocked*, vol. 20, p. 100377, 2020, doi: 10.1016/j.imu.2020.100377.
- [3] A. J. Jenkins, M. V. Joglekar, A. A. Hardikar, A. C. Keech, D. N. O'Neal, and A. S. Januszewski, "Biomarkers in Diabetic Retinopathy," *Rev. Diabet. Stud.*, vol. 12, no. 1–2, pp. 159–195, 2015, doi: 10.1900/RDS.2015.12.159.
- [4] N. Tsiknakis et al., "Deep learning for diabetic retinopathy detection and classification based on fundus images: A review," *Comput. Biol. Med.*, vol. 135, p. 104599, Aug. 2021, doi: 10.1016/j.combiomed.2021.104599.
- [5] H. Safi, S. Safi, A. Hafezi-Moghadam, and H. Ahmadi, "Early detection of diabetic retinopathy," *Surv. Ophthalmol.*, vol. 63, no. 5, pp. 601–608, Sep. 2018, doi: 10.1016/j.survophthal.2018.04.003.
- [6] J. P. Medhi and S. Dandapat, "An effective fovea detection and automatic assessment of diabetic maculopathy in color fundus images," *Comput. Biol. Med.*, vol. 74, pp. 30–44, Jul. 2016, doi: 10.1016/j.combiomed.2016.04.007.
- [7] L. Hill and L. E. Makaroff, "Early detection and timely treatment can prevent or delay diabetic retinopathy," *Diabetes Res. Clin. Pract.*, vol. 120, pp. 241–243, Oct. 2016, doi: 10.1016/j.diabres.2016.09.004.
- [8] J. K. H. Goh, C. Y. Cheung, S. S. Sim, P. C. Tan, G. S. W. Tan, and T. Y. Wong, "Retinal Imaging Techniques for Diabetic Retinopathy Screening," *J. Diabetes Sci. Technol.*, vol. 10, no. 2, pp. 282–294, Mar. 2016, doi: 10.1177/1932296816629491.
- [9] M. W. Nadeem et al., "Brain Tumor Analysis Empowered with Deep Learning: A Review, Taxonomy, and Future Challenges," *Brain Sci.*, vol. 10, no. 2, p. 118, Feb. 2020, doi: 10.3390/brainsci10020118.
- [10] M. Anam et al., "Osteoporosis Prediction for Trabecular Bone using Machine Learning: A Review," *Comput. Mater. Contin.*, vol. 67, no. 1, pp. 89–105, 2021, doi: 10.32604/cmc.2021.013159.
- [11] M. W. Nadeem, H. G. Goh, A. Ali, M. Hussain, M. A. Khan, and V. A. Ponnusamy, "Bone Age Assessment Empowered with Deep Learning: A Survey, Open Research Challenges and Future Directions," *Diagnostics*, vol. 10, no. 10, p. 781, Oct. 2020, doi: 10.3390/diagnostics10100781.
- [12] M. I. Razzak, S. Naz, and A. Zaib, "Deep Learning for Medical Image Processing: Overview, Challenges and the Future," in *Lecture Notes in Computational Vision and Biomechanics*, vol. 26. Cham: Springer International Publishing, 2018, pp. 323–350. doi: 10.1007/978-3-319-65981-7\_12.
- [13] K. Xu, D. Feng, and H. Mi, "Deep Convolutional Neural Network-Based Early Automated Detection of Diabetic Retinopathy Using Fundus Image," *Molecules*, vol. 22, no. 12, p. 2054, Nov. 2017, doi: 10.3390/molecules22122054.
- [14] S. Keel, J. Wu, P. Y. Lee, J. Scheetz, and M. He, "Visualizing Deep Learning Models for the Detection of Referable Diabetic Retinopathy and Glaucoma," *JAMA Ophthalmol.*, vol. 137, no. 3, p. 288, Mar. 2019, doi: 10.1001/jamaophthalmol.2018.6035.
- [15] F. Zabihollahy, A. Lochbihler, and E. Ukwatta, "Deep learning based approach for fully automated detection and segmentation of hard exudate from retinal images," in *Medical Imaging 2019: Biomedical Applications in Molecular, Structural, and Functional Imaging*, B. Gimi and A. Krol, Eds., San Diego, United States: SPIE, Mar. 2019, p. 7. doi: 10.1117/12.2513034.
- [16] S.-I. Pao, H.-Z. Lin, K.-H. Chien, M.-C. Tai, J.-T. Chen, and G.-M. Lin, "Detection of Diabetic Retinopathy Using Bichannel Convolutional Neural Network," *J. Ophthalmol.*, vol. 2020, pp. 1–7, Jun. 2020, doi: 10.1155/2020/9139713.
- [17] V. S and V. R, "A Survey on Diabetic Retinopathy Disease Detection and Classification using Deep Learning Techniques," in *2021 Seventh International conference on Bio Signals, Images, and Instrumentation (ICBSII)*, Chennai, India: IEEE, Mar. 2021, pp. 1–4. doi: 10.1109/ICBSII51839.2021.9445163.
- [18] A. Elhadad, F. Alanazi, A. I. Taloba, and A. Abozeid, "Fog Computing Service in the Healthcare Monitoring System for Managing the Real-Time Notification," *J. Healthc. Eng.*, vol. 2022, pp. 1–11, Mar. 2022, doi: 10.1155/2022/5337733.
- [19] M. Yadav, R. Goel, and D. Rajeswari, "A Deep Learning Based Diabetic Retinopathy Detection from Retinal Images," in *2021 International Conference on Intelligent Technologies (CONIT)*, Hubli, India: IEEE, Jun. 2021, pp. 1–5. doi: 10.1109/CONIT51480.2021.9498502.
- [20] M. T. Al-Antary and Y. Arafa, "Multi-Scale Attention Network for Diabetic Retinopathy Classification," *IEEE Access*, vol. 9, pp. 54190–54200, 2021, doi: 10.1109/ACCESS.2021.3070685.

- [21] W. Kusakunniran et al., "Detecting and staging diabetic retinopathy in retinal images using multi-branch CNN," *Appl. Comput. Inform.*, Dec. 2022, doi: 10.1108/ACI-06-2022-0150.
- [22] C. Raja and L. Balaji, "An Automatic Detection of Blood Vessel in Retinal Images Using Convolution Neural Network for Diabetic Retinopathy Detection," *Pattern Recognit. Image Anal.*, vol. 29, no. 3, pp. 533–545, Jul. 2019, doi: 10.1134/S1054661819030180.
- [23] S. H. Abbood, H. N. A. Hamed, M. S. M. Rahim, A. Rehman, T. Saba, and S. A. Bahaj, "Hybrid Retinal Image Enhancement Algorithm for Diabetic Retinopathy Diagnostic Using Deep Learning Model," *IEEE Access*, vol. 10, pp. 73079–73086, 2022, doi: 10.1109/ACCESS.2022.3189374.
- [24] A. Bilal, L. Zhu, A. Deng, H. Lu, and N. Wu, "AI-Based Automatic Detection and Classification of Diabetic Retinopathy Using U-Net and Deep Learning," *Symmetry*, vol. 14, no. 7, p. 1427, Jul. 2022, doi: 10.3390/sym14071427.
- [25] M. S. B. Phridviraj, R. Bhukya, S. Madugula, A. Manjula, S. Vodithala, and M. S. Waseem, "A bi-directional Long Short-Term Memory-based Diabetic Retinopathy detection model using retinal fundus images," *Healthc. Anal.*, vol. 3, p. 100174, Nov. 2023, doi: 10.1016/j.health.2023.100174.
- [26] K. Loheswaran, "Optimized KFCM Segmentation and RNN Based Classification System for Diabetic Retinopathy Detection," in *ICCCE 2020*, A. Kumar and S. Mozar, Eds., in *Lecture Notes in Electrical Engineering*, vol. 698. Singapore: Springer Nature Singapore, 2021, pp. 1309–1322. doi: 10.1007/978-981-15-7961-5\_119.
- [27] K. V. Spoorthi and B. S. Rekha, "Diabetic Retinopathy Prediction using Deep learning," in *2021 IEEE International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)*, Bangalore, India: IEEE, Dec. 2021, pp. 1–6. doi: 10.1109/CSITSS54238.2021.9683553.
- [28] A. Chakravarty and J. Sivswamy, "A Deep Learning based Joint Segmentation and Classification Framework for Glaucoma Assesment in Retinal Color Fundus Images," 2018, doi: 10.48550/ARXIV.1808.01355.
- [29] T. Liu, Y. Tian, S. Zhao, X. Huang, and Q. Wang, "Automatic Whole Heart Segmentation Using a Two-Stage U-Net Framework and an Adaptive Threshold Window," *IEEE Access*, vol. 7, pp. 83628–83636, 2019, doi: 10.1109/ACCESS.2019.2923318.

# Feline Wolf Net: A Hybrid Lion-Grey Wolf Optimization Deep Learning Model for Ovarian Cancer Detection

Dr. Moresh Mukhedkar<sup>1</sup>, Divya Rohatgi<sup>2</sup>, Dr. Veera Ankalu Vuyyuru<sup>3</sup>, Dr K V S S Ramakrishna<sup>4</sup>,  
Prof. Ts. Dr. Yousef A. Baker El-Ebiary<sup>5</sup>, Dr. V. Antony Asir Daniel, M.B.A., M.E., Ph.D<sup>6</sup>

Assistant Professor, D Y PATIL UNIVERSITY, Pune, India<sup>1</sup>  
Dept. of CSE-ASET, Amity University, Maharashtra, India<sup>2</sup>

Assistant Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation,  
Vaddeswaram, 522502, A.P, India<sup>3</sup>

Department of CSE, Vignan's Nirula Institute of Technology and Science for Women, Pedapalaluru,  
Guntur-522005, Andhra Pradesh, India<sup>4</sup>

Faculty of Informatics and Computing, UniSZA University, Malaysia<sup>5</sup>

Associate Professor and Head of the Department, Department of Electronics and Communication Engineering, Loyola Institute of  
Technology & Science, Kanyakumari-629302, Tamilnadu, India<sup>6</sup>

**Abstract**—Ovarian cancer is a major cause of mortality among gynecological malignancies, emphasizing the critical role of early detection in improving patient outcomes. This paper presents an automated computer-aided design system that combines deep learning techniques with an optimization mechanism for accurate ovarian cancer detection that utilizes pelvic CT images dataset. The key contribution of this work is the development of an optimized Bi-directional Long Short-Term Memory (Bi-LSTM) model which is introduced in the layers of CNN (Convolutional Neural Network), enhancing the learning process. Additionally, a feature selection method based on Lion with Grey Wolf Optimization (LGWO) is employed to enhance classifier efficiency and accuracy. The proposed approach classifies ovarian tumors as benign or malignant using the Bi-LSTM model, evaluated on the Ovarian Cancer University of Kaggle dataset. Results showcase the effectiveness of the method, achieving remarkable performance metrics, including 98% accuracy, 99.7% recall, 93% precision, and an impressive F1 score of 98%. The proposed method's efficiency is validated through comparison with validating data, demonstrating consistent and reliable results. The study's significance lies in its potential to provide an accurate and efficient solution for early ovarian cancer detection. By leveraging deep learning and optimization, the proposed method outperforms existing approaches, highlighting the promise of advanced computational techniques in improving healthcare outcomes. The findings contribute to the field of ovarian cancer detection, emphasizing the value of integrating cutting-edge technologies for effective medical diagnosis.

**Keywords**—Ovarian cancer; deep learning; bidirectional long short term memory; CT images; convolutional neural network; lion grey wolf optimization

## I. INTRODUCTION

Ovarian cancer is the occurrence of irregular cells in the ovary which reproduce uncontrollably and then causes tumor malignancy in the tissues [1]. This can be categorized into three major types namely sex-cord- stromal, epithelial and

germ cell. In this epithelial ovarian cancer has a secondary type namely clear cell, mucinous, serous and endometriosis. Serous tumors are segregated into low-grade serious carcinomas (LGSC) and high-grade serous carcinomas (HGSC). By analysis it shows that 70-80% people cause epithelial cancer, clear cell in addition with mucinous are induced to less than 5% people and 10% people cause endometriosis. The first sign of ovarian cancer is epithelial cancer, which must be detected early to prevent death. Ovarian cancer should be detected in early stage, for detecting involves various methods and this paper proposed the ovarian cancer through deep learning [2]. Epithelial ovarian cancer made up of diverse ancient subsets with peculiar genomic features, they are enhancing the accuracy and effectiveness of healing treatment. It enables the detection of reaction such as ovarian cancer susceptibility genes BRCA2 and BRCA1 and also corresponded combining deficiency with damage in DNA response pathway resistance either inhibitor. This procedure is to process genomic transformation in tumors along with blood to evaluate sensitivity, therapy resistance and precise residual disease indicator. Around 230000 women shall be analyzed and 150000 women are died. This signifies that EOC is the seventh usually detected cancer in women. Genetic syndromes comprise Peutz-Jegher along with uncommon disorders and they are nevoid basal cell sign. Their causative factor includes an embryonic pregnancy, initial menarche, fallen age menopause, smoking, polycystic ovary syndrome [3].

Ovarian cancers are predicted by two medical examinations they are serum cancer antigen 125(CA125) and transvaginal ultrasound. Specificity along with sensitivity is restrictions. CA125 is frequently raised in benign stage including endometriosis and ovarian cysts. It is not determined in initial stage. This test evaluates the quantity of protein named CA-125 within the blood. High stages of CA-125 are occurred in women. This test is beneficial as tumor marker to identify as ovarian cancer. High level of CA125 persons fails



to validate ovarian cancer because it might be endometriosis and pelvic inflammatory disease. Person with abnormal CA-125 level, medical practitioner again examines the report. CA-125 is widespread in countries such as Canada, Australia, United States and Ireland as an prediction for ovarian cancer [4]. Transvaginal Ultrasound (TVUS) is a checkup that utilizes sound waves to monitor the uterus, fallopian tubes and ovaries by fixing ultrasound wand inside the vagina. It detects massive tumor or benign take placed [5]. The most present worldwide statistic calculates 295,414 recently predicted cases of ovarian cancer all year and also this disease causes 184799 yearly dead. Moreover, in an advanced phase 75% of ovarian cancer patients are difficult to diagnose properly. The treatment of chemo resistant ovarian cancer by targeting mitochondria has observed. It is developed under the terms where cancer cells modify and swap to mitochondrial respiration this turns severe to endurance so that perfect metabolic focus for chemo resistant ovarian cancer. In ovarian cancer tumor metastasis together with chemoresistance are allied in mitochondria spatial redistribution. Aberrant dependence of mitochondrial pathways are caused by specific sets of genetic mutation [6].

Recent treatment strategies include by combining debulking surgery, radiation therapy and drug treatment. Some sophisticated remedies are immunotherapy, hormone therapy and targeted therapy. OC treatment mainly involves chemotherapy. Repeatedly applied chemotherapeutic agents employed for curing of ovarian cancer includes platinum comprising drugs i.e., carboplatin along with cisplatin and tiane family i.e., docetaxel together with paclitaxel. In recent days, scientists have been providing an immunotherapy approach to treat gynecologic cancers. From ascites, tumor and blood of ovarian cancer patients the T cells and Antibodies are predicted in immunotherapy [7]. The machine learning algorithms are involved in several data mining applications which include ovarian cancer. Although these algorithms don't perform suitably due to inaccurate computational complexity, data imbalance, and missing values. Recently deep learning-based process has gained advantage in computer vision related applications and data mining over machine learning. Deep Learning Network utilized in several regions such as speech recognition, computer vision, healthcare, image processing [8]. In this research, emphasize on developing a deep learning established system for classification of ovarian cancer. The systematic study highlights the Bi-LSTM classifier which effectively worked and maintained more accuracy than other classifier such us CNN, 3D-CNN, KNN and Random Forest classifier. The deep learning handles the data by various phases including convolution layer, dense layer and pooling layer. Commonly deep learning have sequential architecture pursued by feature extraction, feature selection and classification [9]. Deep learning is frequently used invading detection of ovarian cancer. Deep learning scheme Bi-LSTM network performs a vital function to enhance the classification [10]. Optimization algorithms can be used to automatically identify and extract important features from the data in the image especially size, shape, and area of the tumor occur in ovary. This helps to remit the time required and potential for manual analysis, and can also develop the accuracy of the detection process.

Optimization algorithms can also be used to train deep learning models that can automatically classify medical data as either containing a tumor or not. These models can be optimized to improve their performance, such as by adjusting their parameters or optimizing their training process.

The key contribution of the described outline is given as follows:

- The development of an optimized Bi-LSTM model for the early detection of ovarian cancer.
- The novelty of this work lies in the development of an optimized Bi-LSTM model for ovarian cancer detection, which enhances the learning process and achieves impressive performance metrics.
- By utilizing deep learning techniques and an optimization mechanism, this model enhances the learning process and improves the accuracy of classification.
- Additionally, a feature selection method called Lion with Grey Wolf Optimization (LGWO) is employed to further enhance the efficiency and effectiveness of the classifiers.
- The combination of the Bi-LSTM model and the LGWO feature selection method allows the system to accurately classify ovarian tumors as either benign or malignant.
- To validate the efficiency and reliability of planned method, its presentation metrics are compared with the validating data, demonstrating consistent and reliable results.

The configuration of this essay is as accompanies: Section II contains the related work that is framed to understand the proposed paper with the existing methods while Section III elaborates the problem statement. Section IV depicts the proposed LGWO-Bi-LSTM architectures. The results and performance metrics are tabulated and graphically represented in Section V. Finally, in Section VI, conclusion and future works are presented.

## II. RELATED WORK

Srivastava et al. [11] outlined adjusted VGG-16 deep learning, ovarian cyst detection was accomplished approach. Pretrained deep learning method consists of VGG-16 model. To understand the VGG-16 model keenly a multiple of 3×3 kernel sized filters are obtained. This explains the established VGG-16 prototype fine-tuned with the data-set of ultrasonic images. Modifying the VGG-16 model's top four layers allows for fine-tuning. Distinct women's sample ovarian images are gathered to detect if ovarian cyst is occurred or not. The accuracy of 92.11% is obtained. The downside of VGG-16 deep learning network is slower than the ResNet architecture. Meng et al. [12] Intended virtual historical method of staining with deep generative adversarial mechanism to determine the ovarian cancer. Here hematoxylin and also Eosin staining method are used. Depending upon GAN, here autofluorescence images are created to develop a weakly supervised learning method of unstained ovarian tissue

regions depending upon E and H staining portion of ovarian tissues. The characteristic of the result predicted by the approach is more precise. This paper proposed a valid autofluorescence image generation of algorithm in the insufficient data which can consume time and laborious data representations in many situations. Through doctors determined the accurate unstained fluorescence image of ovarian cancer can be generated by this mechanism is 93%. The existing algorithm involves several complexities to predict the ovarian cancer. The visual evaluation established through deep learning method reaches an accuracy of 95%. Lu et al.[13] Proposed the determination of ovarian cancer by machine learning. It detects the exact of benign and ovarian tumors. In this method includes several processes like blood test, demographics, tumor markers and general chemistry. 349 Chinese individuals' information regarding 49 characteristics were obtained, and 235 patients' data were processed using the machine learning Minimum Redundancy - Highest Relevancy approach. The MRMR adopts 10 noteworthy characteristics, including human epididymis protein 4 (HE4) and cancer-embryonic antigen (CEA) are top featured by the decision tree model. It shows that machine learning is mostly incredible in detecting the complex diseases. The accuracy of trained data of ROMA, Decision Tree, and Logistic Regression is 79.6%, 87.2%, 84.7% and the accuracy of test data of ROMA, Decision Tree, and Logistic Regression is 92.1%, 95.6%, 97.4% respectively. This method is largely unstable contrasted with the other decision approaches. Schwartz et al.[14] Proposed an automated structure to recognize ovarian cancer in transgenic mice through optical coherence tomography (OCT) recordings. Classification is achieved by a neural network that distinguishes structurally ordered sequence of tomograms. It consists of three neural networks they are 3D convolutional neural network, VGG-supported feed-forward network and convolutional long short-term memory network. The outcome indicates that it can accurately output manual tuning although there is a default in noise inherent OCT images. It obtained a mean of  $0.81 \pm 0.037$ .

Yang et al. [15] Proposed the detection of ovarian cancer by combining the clinical significance of salivary mRNAs together with carcinoembryonic antigen. It is a liquid biopsy method and it can be used to predict many types of cancers. Here we determined a discriminatory study of ovarian cancer by uniting CEA and salivary mRNA biomarkers. Using two methods this technique is achieved. They are independent validation phase and discovery phase in which finding and evaluating of multiple biomarkers is enabled by discovery phase, independent validation phase confirms the availability of the finest bio-markers. Discovery phase categories blood's CEA level and five mRNA biomarkers in saliva are noted. Novel panel of biomarkers is used to segregate ovarian cancer patients and healthy people with high specificity of 82.9% and also with high sensitivity of 89.3% are obtained. In validation phase it acquires sensitivity 85% and simplicity 88.3%. Lemmings et al. The author in [6] proposed the treatment of chemo resistant ovarian cancer by targeting mitochondria. It is developed under the terms where cancer cells modify and swap to mitochondrial respiration this turns severe to endurance so that perfect metabolic focus for chemo resistant ovarian cancer. In ovarian cancer tumor metastasis together

with chemoresistance are allied in mitochondria spatial redistribution. Zhang et al. [16] For manufacturers to offer more advanced and inexpensive, high-quality Ultrasonic Flaw (UT) technology that could improve maternity medical services, it is recommended that we tackle the requirement to establish sustained sonar criteria with tolerably high pregnant and neonatal death rates. A group of artificial intelligence techniques, like the most recent advances in training to learning for maternal ultrasonography, is growing in acceptance and sparking excitement across a number of industries, such processing images for artificial intelligence. In this study, sophisticated artificial intelligence (AI) algorithms that use logistic reconstruction classifiers (LRC) and convolution neural networks (CNNs) to associate the original input picture to the intended outcome image are of special interest. Additionally, they employed the Internet of Medical Things (IoMT) to divide up maternal tumour imaging and identify tumours for specialists. The experimental findings demonstrate that the LRC based on CNN may be utilised for predicting the outcomes of maternity ultrasounds with enhanced levels of paternal and neonatal movement.

Kaggie et al. [17] proposed Ovarian cancer is one of several tumour subgroups that may be characterized using multiparametric magnetic resonance imaging (MRI). Despite the limitations of traditional anecdotal T1- and T2-weighted scans, quantified mappings of MRI relaxation values, including T1 and T2 the mapping process, shows promise for improved tumour evaluation. Nevertheless, due to the ordered measurement of several parameters, quantitative MRI relaxation mapping approaches sometimes need lengthy scan periods. Fast qualitative MRI is made possible by a novel technique called magnetic resonance fingerprinting (MRF), which takes use of transitory signals brought on by changes in the parameters that make up of a pseudorandom programme. Statistical correspondences are subsequently created by matching these temporary signals to a computed lexicon of T1 and T2 elements. The capacity of the MRF to monitor several factors concurrently may provide a novel method for identifying malignancy and evaluating the effectiveness of treatments. Using ovarian cancer as an illustration structure, this practical research examines MRF for concurrent T1, T2, and relative proton density (rPD) mapping. Negi et al. [18] proposed In comparison to monolayer OLED design, a unique three hole block layer (HBL) arrangement of the OLED is presented that exhibits improved luminosity of 25285 cd/m<sup>2</sup>. The 74% increase in illumination power efficiency is partly to blame. The internal circuit evaluation is used to confirm an extensive numerical analysis built on the Poisson and drift dispersion model. The results of the examination show that the suggested gadget has a higher rate of recombine. A higher amount of recombination is a result of effective whole shielding as well as elevated particle input. Consequently, ovarian cancer is diagnosed by triple HBL OLED. The gadget produced an ultimate photon current level of 93 mA and had good responsiveness to different bands. According on their urine's ability to shine, a healthy individual can be distinguished from an oncological cancer survivor.

### III. PROBLEM STATEMENT

The detection of ovarian cancer is crucial for preventing patient mortality. However, there are several challenges associated with this process, such as the risk of infection, bleeding, blood clot formation, swollen legs, fatigue, infertility, and bowel changes. To aid in the detection process, a fine-tuned VGG-16 model has been developed to categorize ovarian cysts based on their forms, cysts, HOC, dermoid cysts, and PCOS. However, the current system lacks a user interface that allows pathologists to conveniently upload and predict images, which needs to be addressed and improved [19]. The

fine-tuned VGG deep learning network more time to train its parameters [11].

### IV. PROPOSED LGWO-BASED BI-LSTM

Initially the ovarian cancer dataset is preprocessed through weaned filter and then image segmentation is employed by Fuzzy C-means clustering. In this way feature is extracted using Grey Level Co-occurrence Matrix and feature is adopted by hybrid Lion Grey Wolf optimization and also categorized by Bi-LSTM model. Block diagram of hybrid Lion Grey Wolf Optimization through Bi-LSTM model is depicted in Fig. 1.

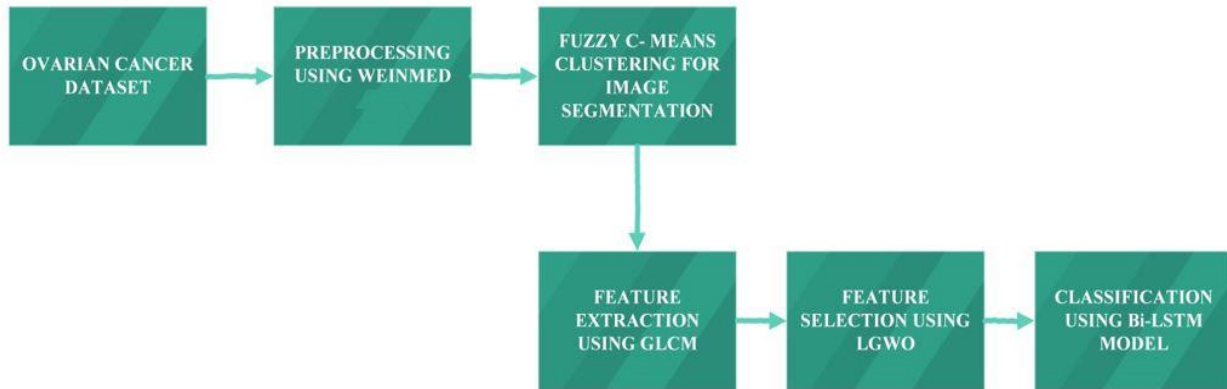


Fig. 1. Block diagram of hybrid lion grey wolf optimization through Bi-LSTM model.

#### A. Data Collection

The research utilizes pelvic CT image datasets affiliated with China's Qingdao University, specifically from a Class 3 hospital. After screening the obscured data, it acquired a totality of about of 5100 CT images of 223 patients containing pelvic CT images with highlights were gathered and utilized. The images were divided into two equal sets, with 50% used as training data and the other 50% used as testing data. This means that 2550 images with highlights were utilized for training and 2550 CT images with highlights were utilized for testing. The dataset consists of 2000 pelvic CT images collected from 223 patients. It contains three types of pelvic CT images containing clear cell, mucinous, endometriosis and serous. Pelvic CT images taken for mucinous, clear cell and endometriosis is 2000, 1500, and 1600 for ovarian cancer [20].

#### B. Pre-processed by Weinmed Filter

The most crucial stage for the best categorization outcomes is pre-processing. It is frequently carried out on data before classification to make sure the desired outcomes are reached [21]. The CT images that obtained are pre-processed to get more relevant for further process. The preprocessing step includes smoothing, toughen the edges of the image and noise removal. The objective of pre-processing enhances the image quality. Each pixel replaced weinmed filter with average value of intensities in region. Median filter removes salt and noise accurately. Weinmed filter probes to construct an optimal determination of the original image by enforcing the minimum mean square error limit between estimate and original image. The purpose of the weinmed filter is to reduce

the mean square error and it also de-blurs the image and reverses filtering. This filter also opens morphological operation and contrast enhancement to image enhancement and reduce noise. Individual background noises calculate the contrast of each region. By preserving the edge information of unclear areas, the background noise is removed shall highlights the digital prototype. Both the degradation function and noise are managed by weinmed filter. Through the degradation model, the error among the estimated signal and the input signal is provided.

Weiner filter formula is expressed in Eq. (1),

$$F_{X(u,v)} = \frac{|H(u,v)|^2}{|H(u,v)|^2 + S \frac{P_n(u,v)}{P_f(u,v)}} G(u,v) * \frac{1}{H(u,v)} \quad (1)$$

$P_n(u,v)$  - Noise power spectrum

$P_f(u,v)$  - Power Spectrum of the original image

$H(u,v)$  -Fourier Transform of the point spread function

$G(u,v)$  -Power Spectrum with the Fourier change of the noisy point cloud image

$F_{X(u,v)}$  -Result Value After restoration

The computed de-noising equation is given below.

Let FUV stand for the collection image parameters in a k-by-1-inch square sub image windows centered at the two points (u, v). The following phase in preparation is to remove and normalize the surrounding data during processing itself, albeit this might affect the results of segmentation. During this

moment, the edge from the CT ovarian equation is identified using a clever edge detecting approach Eq. (2) [22].

$$f(u, v) = \text{Median}\{F_{X(u,v)}\} \in Fu, v \quad (2)$$

The above equation (2) is termed as Weinmed filter.

### C. Segmentation using Fuzzy C-means Clustering

The segmentation process involves identifying and labeling the different regions in the CT images, such as the mucinous, clear cell, endometriosis, and serous. This task can be challenging due to the complex and heterogeneous nature of ovarian cancer, which can exhibit variations in size, shape, location, and intensity. In addition, there can be variability in the imaging protocols and noise artifacts that can affect the quality of the images. Segmentation of ovarian cancer using CT images can be performed using a variety of techniques, such as threshold, region-growing, active contours, and machine learning-based approaches. Here, Fuzzy clustering method is adopted to perform the task. In the context of ovarian cancer detection, image segmentation using fuzzy clustering can be used to separate the tumor region from the ovarian cancer. The segmentation result can then be used for quantitative analysis and clinical decision-making. Several studies have reported promising results using fuzzy clustering for ovarian cancer segmentation.

Fuzzy C-means clustering is preferable optimization for feature segmenting the CT images. It is the most common fuzzy algorithm which is extremely delicate to outliers, noise, and size of the clusters. It starts by randomly initializing the cluster centroids and the fuzziness parameter, and then iteratively updates the centroids and the degree of membership until convergence. The segmentation result is obtained by assigning each pixel to cluster with highest degree of membership. After initializing degree of membership values, the cluster centroids are calculated and then according on how far there is among every point of data to the geographic center of every cluster, the range of membership metrics associated with each observation point was modified. The levels of values for membership have been updated the cluster centers are re-computed using the weighted average of the data points. Repeat the process until convergence. Once, the algorithm converges, the final assignment of each data point to a cluster is based on the highest degree of membership value [23].

Let us consider a number of finite dataset be 'n' and A = (a<sub>1</sub>, a<sub>2</sub>, ..., a<sub>n</sub>). The dataset A is divided in to group of cluster by using FCM algorithm. Formula for FCM algorithm is given in Eq. (3).

$$J_s = \sum_{p=1}^n \sum_{q=1}^c u_{pq}^s \|A_p - C_q\|^2 \quad (3)$$

Where,  $u_{pq} = \frac{1}{\sum_{k=1}^c \left(\frac{d_{pq}}{d_{kq}}\right)^{\frac{2}{s-1}}}$  is the degree of membership

value of  $x_p$  in  $q^{th}$  cluster,  $x_p$  is  $p^{th}$  data points and  $c_q = \frac{\sum_{q=1}^n u_{pq}^s x_p}{\sum_{q=1}^n u_{pq}^s}$  is  $q^{th}$  cluster center, s denotes the fuzzy parameter and Euclidean distance is represented as  $\| \cdot \|$ .

Where,  $d_{pq}$  indicates the aloofness among data points  $x_p$  and  $q^{th}$  cluster focus and  $d_{kq}^*$  is distance between  $k^{th}$  cluster focus and  $q^{th}$  cluster center.

### D. Feature Extraction using Scale-Invariant Feature Transform (SIFT)

Finding the initial corresponding characteristics is the aim of the initial phase. The starting imagery A and those perceived image B, both of whom are fed to SIFT to extract and characterize local characteristics, are two separate images that need to be registered  $F_A$  and  $F_B$ , accordingly.

The prospective features are found by scanning across all of the available sizes and places of residence for extracted features. The regional maximal and minimal values of E(x), which can be described as the combination of the function using an image, i.e.,  $F_A$  or  $F_B$ , are used to discover scale-space extreme. Another thorough match to the neighboring data points is carried out to determine the precise positions of the characteristics [24]. Some prospective feature points have inadequate contrast elements or edge locations with poor localization, thus the contrasting values and primary curved proportions are utilized to exclude the unstable characteristics. After reliable prospective characteristics are discovered, every key point is given a dominant orientations determined by the current picture gradients directions histogram to achieve image rotational consistency [25]. A description matrix that is calculated for every distinctive point comes next. A 3-D histogram of grade dimensions and directions serves as such a description. To create a 128-D descriptor, the gradients alignment angle is compressed into eight orientations increments and the placement is compressed into a 4 4 placement grid.

$$\theta = \cos^{-1} \left( \frac{Z_p^H Y_p}{\|Z_p\| \|Y_p\|} \right) \quad (4)$$

where,  $Z_p$  represents the p th feature descriptors vector in that the sensed imagery,  $Y_p$  represents the p th feature descriptor vector in the reference image, and represents the angle among the two variables. The primary maximum degree versus the second minimum angle ratio, represented by the symbol proportion, is utilized to increase the accuracy of matching the initial characteristics. Initial match characteristics are disregarded if their angle ratios of distance are higher than a threshold value.

The combination of entropy together with the mean value of Fourier co-efficient is obtained to provide feature vector in AGLCM algorithm.

### E. Feature Selection using Hybrid LGWO

The Lion with Grey Wolf Optimization algorithm is employed to reducing the errors in training data and detects the malignant portion. Here the leader may be male or female indicated as alpha which enables hunting, sleeping, time to wake, location. Alpha makes decision and Beta handles.

Omega is the lower section of the grey wolf and for further wolves on every occasion [26]. Omega ( $\omega$ ) is noticed as the leftover provision [27]. In Lion optimization the parameters include learning rate ( $\alpha$ ), weight values ( $w$ ), number of layers ( $L$ ), kernel sizes ( $k$ ) and dropout rate ( $\gamma$ ).

1) *Initialization process*: At first the preprocessed output data is initialized as a, B and C as coefficient vectors.

Fitness evaluation:

Eq. (5) evaluates the fitness utility and predicts the result

$$Fit_i = \max accuracy \quad (5)$$

Differentiate the solution depend on filters:

Let the initial fitness be  $d_\alpha$ , the second finest fitness be  $d_\beta$ , the third finest fitness be  $d_\delta$

a) *Roaming*: In this process the territory is visited by every male lion where the total location to number of this location is %S. When the Lion find the best location it reforms the location during roaming. Lion shifts to improved location is denoted in Eq. (6) [28]

$$\gamma \sim U(0,2) \times g \quad (6)$$

Where,  $g$  – Male lion' current location and selected location from the territory and  $U$  – Uniform Distribution

$P_{si} = 0.1 + \min\left(0.5, \frac{Y_i - Y_{best}}{Y}\right)$ ,  $i=1, 2, 3 \dots$ no. of nomad lions

Where,  $P_{si}$  = the chance detected for seperation of all nomad lions  
 $Y_i = i_{th}$  Lion's fitness value and  $Y_{best}$  = Fitness value of the best nomad lion.

b) *Encircling prey*: Together with these three contenders by  $\alpha, \beta, \delta$  and also  $\omega$ . In addition for the group to track a prey is by surrounding their location. Encircling or trapping behaviour of grey wolves for pray during hunting is computed using the following Eq. (7) and (8) [29].

$$d(t+1) = d_{ps}(t) = \vec{B} \cdot \vec{K} \quad (7)$$

$$\vec{K} = |\vec{C} \cdot d_{ps}(t+1) - d_{ps}(t)| \quad (8)$$

Where,  $\vec{B} = 2\vec{c}r_1 - \vec{c}$  and  $\vec{C} = 2r_2$ , and  $d_{ps}(t)$  - The prey position,  $B$  and  $C$  – The coefficient vector,  $\vec{c}$  - Linearly lowered from 2 to 0,  $r_1$  and  $r_2$ - Random vector[0,1], and  $t$  - The iteration n numbers.

c) *Hunting*: Hunting is done based upon Lion Optimization. In every pride the female concentrates on a prey in a cluster to feed their pride. To surround the victim and to preserve pride it follows certain mechanisms. All lions fine-tuned their localities to rely on particular locality and the group associates' localities. Due to this reason the seekers surround the victim and apply relative training, assaults from locality conflict. For this the seekers are classified into three subdivisions. In the centroid of seekers, a prey is engaged. During hunting the hunters are chooses in sequence, a false victim is assaulted by decided seekers in order with the crowd

which decides lion to belong it. If a searcher enhances its fitness a false victim escape from searcher and new locality of prey is detected in Eq. (9) [30].

$$d_{ps}' = p + \text{ran}(0,1) \times PI \times (p - h) \quad (9)$$

Hunter represents the present locality seeker who hit to victim and PI denotes the objective seekers increasing rate in Eq. (10).

$$h' = \begin{cases} \text{rand}((2 \times p - h), p) & \text{if } (2 \times p - h) < ps \\ \text{rand}(p, (2 \times p - h)) & \text{if } (2 \times p - h) > ps \end{cases} \quad (10)$$

The new localities of centroid seekers are given bin Eq. (11)

$$h' = \begin{cases} \text{rand}(h, p) & \text{if } (2 \times p - h) < ps \\ \text{rand}(p, h) & \text{if } (2 \times p - h) > ps \end{cases} \quad (11)$$

The above equation produces an arbitrary value from b to c which is upper and lower limits, respectively.

Understanding and improving the effectiveness of the suggested model depends on research into how the HLGWO's characteristics affect the identification of ovarian cancer. These parameters include a wide range of factors, including population size, convergence standards, crossover and mutation rates, and trade-offs between exploitation and exploration. It is feasible to fine-tune the HLGWO algorithm to increase its efficiency and efficacy in optimizing the deep learning model for ovarian cancer diagnosis by carefully modifying and analyzing these parameters. The model's convergence rate and solution quality can be considerably impacted by striking the correct balance between exploration and exploitation. Additionally, preventing premature convergence and finding the most illuminating characteristics for precise identification may be achieved by optimizing parameters like mutation rates and crossover rates. To further improve HLGWO's potential as a diagnostic tool for ovarian cancer, it is crucial to refine and adapt the model for practical clinical applications. This requires understanding how these parameter adjustments influence the model's sensitivity, specificity, and overall accuracy.

#### F. Classification using Bi-LSTM

Classifications of RNN include Bi-LSTM it is enabled for processing of natural language like text classification. The context is captured as it is a powerful model and words depend on sequence [31]. In ovarian cancer classification, Bi-LSTM networks can be used to predict sequences of MRI or CT scans to classify ovarian cancer as benign or malignant. The input sequence can consist of images from different angles, slices, and time points and the network can learn to recognize patterns and features that are indicative of specific ovarian cancer types. Bi-LSTM model is represented in Fig. 2. In addition, Bi-LSTM networks can also be concatenated with other DL techniques, such as CNNs, to enhance the accuracy of the classification. Convolution neural networks are commonly used to extract data from medical images, while Bi-LSTM networks can learn the temporal dependencies between these features. Overall, Bi-LSTM networks are a promising tool for brain tumor classification because they can model the complex relationships and patterns in sequential

data, thus improving the accuracy of classification by utilizing backward hidden states, which contributes to the enhancement of LSTM network learning [32].

Four gates in LSTM neural network are represented in Eq. (12) to (15).

$$c_t = \sigma(M_f x_t + E_f h_{t-1} + c_c) \quad (12)$$

$$d_t = \tanh(M_g x_t + E_g h_{t-1} + c_d) \quad (13)$$

$$e_t = \sigma(M_i x_t + E_o h_{t-1} + c_e) \quad (14)$$

$$f_t = \sigma(M_o x_t + E_o h_{t-1} + c_f) \quad (15)$$

where, Ef, Eg, Ei, Eo represents the weight matrices of the preceding short-term state ht-1. Mf, Mg, Mi, Mo represents the weight matrices of the present input state xt, and cd, ce, cf,

and co are the bias terms. Where, pt-1 represents the previous long-term state.

The present long-term state of the network gt can be evaluated by using Eq. (16), and yt can be evaluated using Eq. (17).

$$g_t = f_t * p_{t-1} + e_t * d_t \quad (16)$$

$$y_t = h_t = f_t * \tanh(g_t) \quad (17)$$

The classification architecture is depicted in below by given Fig. 2. Similar to this, the LSTM layer gains knowledge of the dependencies among various time steps in sequence data.

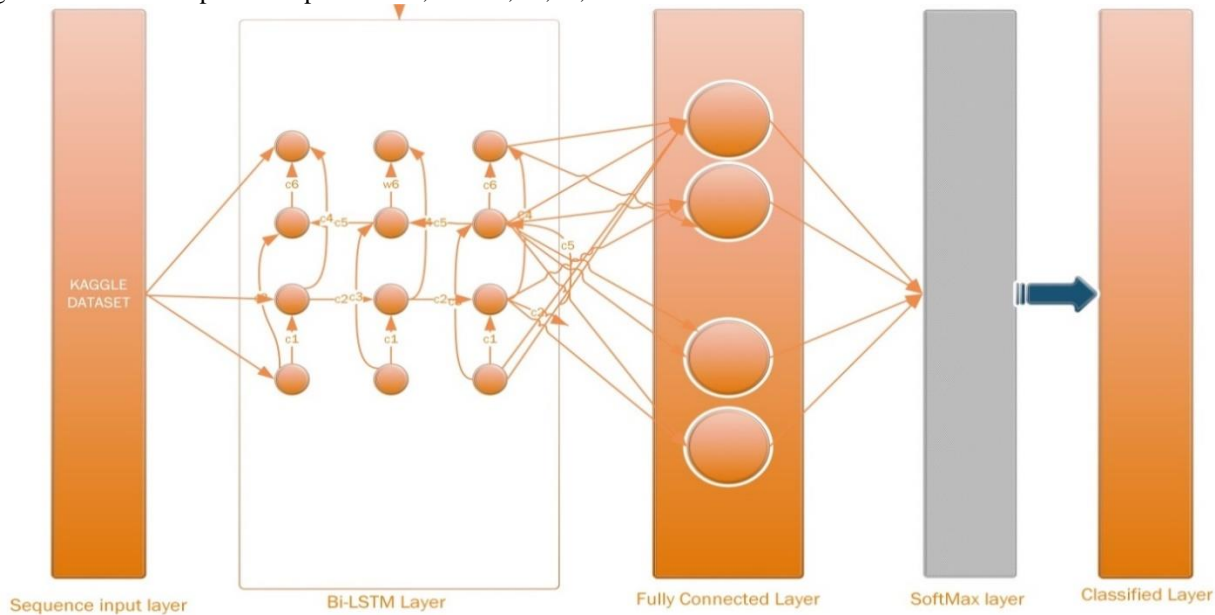


Fig. 2. Bi-LSTM classification of architecture.

**Algorithm 1: LGWO- Bi-LSTM mechanism**

INPUT: CT images from Kaggle dataset

OUTPUT: Classifying the type of ovarian cancer (clear cell, mucinous, serous and endometriosis)

Load input image data

$C = \{c_1, c_2, c_3, \dots, c_n\}$  //Image acquisition

Pre-processing of images

Toughen the edges of the image //Weiner filter

Removes the salt and noise effectively //Median filter

Segmentation of images

Choose the initial cluster and membership matrix //Fuzzy C-means Clustering

Calculate new cluster and new membership matrix

If (Difference in cluster center)

< threshold

Stop

Else

>threshold

Repeat

Feature Extraction

Feature Selection //LGWO

Calculate roaming using the fitness of Lion

Calculate encircling using the fitness of Grey Wolf

Calculate hunting using the fitness of Lion

Classification //Bi-LSTM

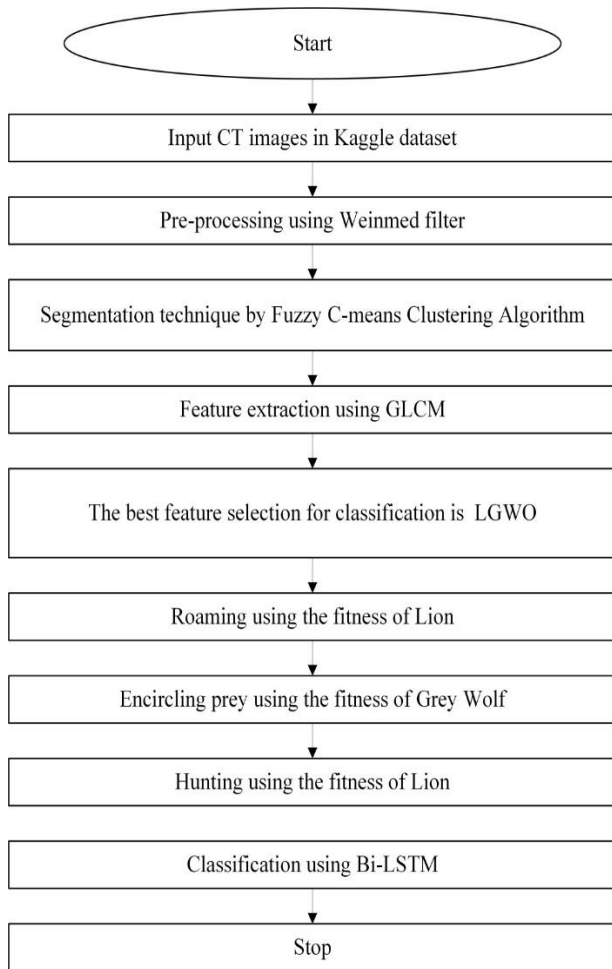


Fig. 3. Flow diagram of HLGWO -BI-LSTM model.

The above Fig. 3 elaborates the overall work flow of the entire proposed model. Main steps involved in detecting and classifying the ovarian cancer shown in flow with algorithmic conditions. This flow chart helps for quick understanding of the presented research.

### V. RESULT AND DISCUSSION

This study provides a detailed explanation of the experimental results obtained from the proposed method, using numerical terms to quantify the outcomes. The validation process was conducted on the CT images. The Kaggle dataset comprising 223 patient’s datasets containing 5100 CT images were separated 50% as testing along with training data. The proposed system’s evaluation was compared with three classification methods: ML-CNN, HOG-ANN, and UBOC. Performance evaluation of the Bi-LSTM classifier and LGW optimization algorithm is carried out using four metrics such as accuracy, F1 score, precision, recall. Entire experimental process is implemented by using MATLAB software in windows 10 platform.

#### A. Evaluation of Performance Metrics

The experiment assessed the models using four evaluation metrics: accuracy, F1-score, precision, and recall. These parameters are specifically defined in Eq. (18) to (21).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (18)$$

$$Recall = \frac{TP}{TP+FN} \quad (19)$$

$$Precision = \frac{TP}{TP+FP} \quad (20)$$

$$F1score = \frac{2*Recall*Precision}{Recall+precision} \quad (21)$$

In above equations, *TP* refers to the no. of data finely classified as positive out of all the data that were actually positive. *TN* Refers to the no. of data finely classified as negative out of all the data that were actually negative. *FN* is the number of data that were mistakenly classified as negative by the model even though they were actually positive in the dataset. *FP* is the number of data that were mistakenly classified as positive by the model even though they were actually negative in the dataset, Recall is defined as the ratio of the no. of data classified as positive by the model to the actual no. of data that were positive in the dataset. Precision is the ratio between no. of data that were correctly classified as positive by the model and the total no. of data defines positive. Finally, F1-score is the harmonic mean of recall and precision, as explained in [33].

TABLE I. EXPERIMENTAL RESULT ANALYSIS FOR DIFFERENT PARAMETERS WITH OTHER METRICS

Method	Accuracy	Recall	Precision	F1-score
ML-CNN[34]	96.5	96	98	97
CNN-DenseNet[35]	94.73	98.9	91	95
AlexNet[36]	84.45	90	75	87
Proposed Bi-LSTM-LGWO	98	99.7	93	98

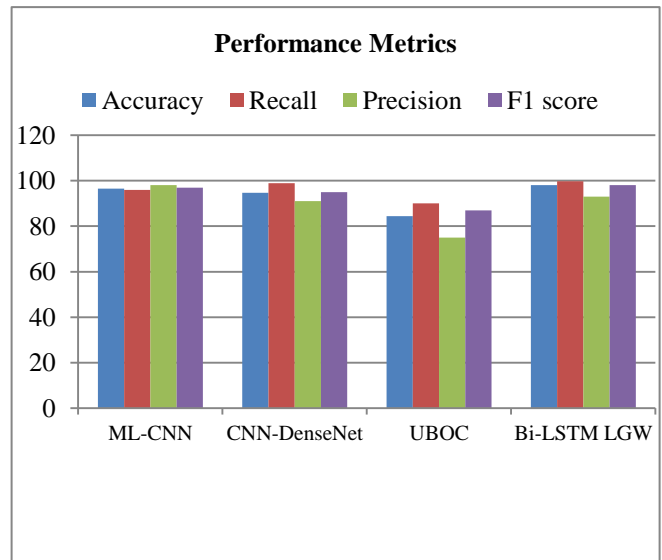


Fig. 4. Performance evaluation of various methods of classification.

In Table I, the performance evaluation of the proposed system is tabulated. The proposed Bi-LSTM-FFO shows higher accuracy when compared with other classifiers. Average of Precision and recall of the existing methods is higher than the proposed method. Mean value of precision

and recall gives a significant measure of classification called F1-score; the graphical identification of performance analysis is shown in Fig. 4.

TABLE II. COMPARED RESULT IN TERMS OF SENSITIVITY AND SPECIFICITY

Method	Specificity	Sensitivity
ML-CNN[34]	89.7	90.57
CNN-DenseNet[35]	96.94	95
AlexNet[36]	95	72
Proposed Bi-LSTM-LGWO	96.4	98

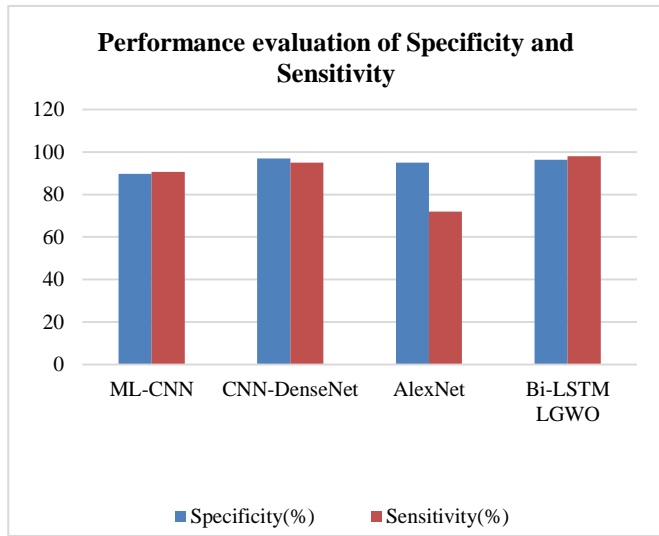


Fig. 5. Comparison of sensitivity and specificity for existing and proposed methods.

Comparison of the proposed model performance results with the existing methods is mentioned in Table II. For clear understanding comparative analysis of sensitivity and specificity are graphically represented in Fig. 5.

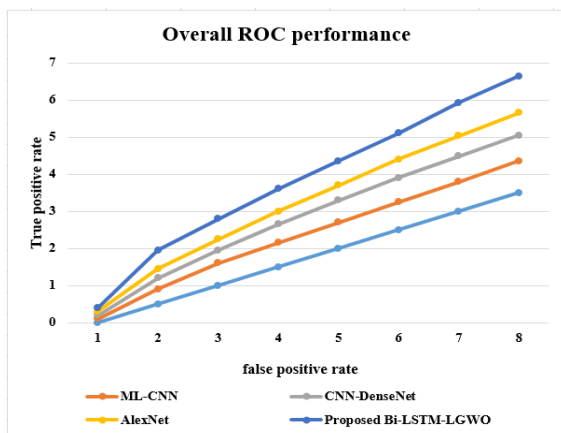


Fig. 6. ROC comparison.

Fig. 6 displays the ROC curve for different models used in ovarian cancer detection. The x-axis represents the FPR, and the y-axis represents the TPR. The proposed Bi-LSTM-LGWO model achieves the highest TPR of 0.7, outperforming

ML-CNN, CNN-DenseNet, and AlexNet, which have TPR values of 0.4 and 0.3.

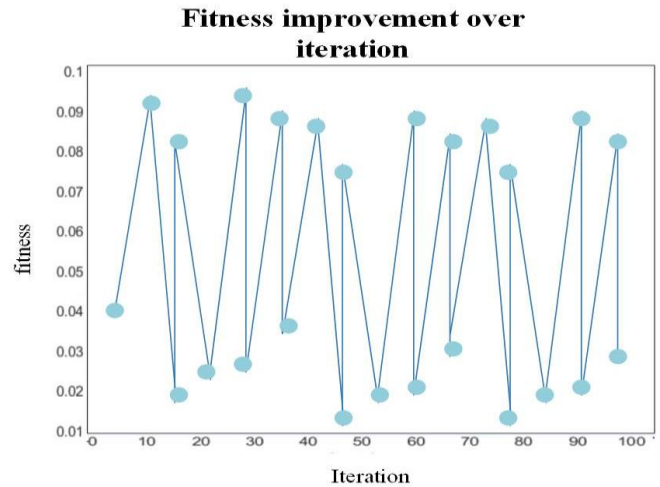


Fig. 7. Fitness improvement.

Fig. 7 shows optimization, fitness improvement refers to the enhancement of the objective function's value or performance metric being optimized. This can be achieved by applying various optimization techniques, such as evolutionary algorithms, gradient descent, or simulated annealing, to iteratively search for better solutions.

**B. Discussion**

The findings of the suggested Bi-LSTM-LGWO model are highly encouraging, with a great sensitivity of 99.7% and an amazing accuracy of 98%. These results indicate that the model performs very well in detecting ovarian cancer properly, which is essential for early diagnosis and management. However, it's important to recognize that the suggested task has certain limits. First off, as larger and more diverse datasets are frequently needed for reliable performance, this may have an impact on how generalizable the model is. Second, the Bi-LSTM-LGWO model's computational complexity could make it impractical to use in real-time clinical situations. Future research should focus on overcoming these constraints by enlarging the dataset and streamlining the model. The model's credibility and suitability for use in clinical practice would also be further increased by investigating new performance indicators and carrying out external validation on various datasets.

**VI. CONCLUSION**

The suggested method is more effective than current classifiers and provides a better level of accuracy, making it a potential direction for further study. With accuracy of 98%, recall of 99.7%, precision of 93%, and F1-measure of 98%, the provided model exceeds the existing model. In comparison to the MNN-CNN method, the adopted classifier in the suggested model is more effective. In this approach Bi-LSTM model was employed for classification. The suggested methodology makes a significant addition to the field of ovarian cancer identification and classification overall. The suggested approach will then be examined by running tests on data sets from various sources and industries. Exploring



various topologies, optimization methods, and hyper parameters can improve the Bi-LSTM classifier and LGWO optimization algorithm's performance. In future proposed methods, performance should be validated through larger-scale clinical trials and collaborations with healthcare professionals to assess its reliability and effectiveness in practical medical settings.

#### REFERENCES

- [1] S. Akter et al., "Recent Advances in Ovarian Cancer: Therapeutic Strategies, Potential Biomarkers, and Technological Improvements," *Cells*, vol. 11, no. 4, p. 650, Feb. 2022, doi: 10.3390/cells11040650.
- [2] C. Stewart, C. Ralyea, and S. Lockwood, "Ovarian Cancer: An Integrated Review," *Semin. Oncol. Nurs.*, vol. 35, no. 2, pp. 151–156, Apr. 2019, doi: 10.1016/j.soncn.2019.02.001.
- [3] S. Lheureux, C. Gourley, I. Vergote, and A. M. Oza, "Epithelial ovarian cancer," *The Lancet*, vol. 393, no. 10177, pp. 1240–1253, Mar. 2019, doi: 10.1016/S0140-6736(18)32552-2.
- [4] "The diagnostic performance of CA125 for the detection of ovarian and non-ovarian cancer in primary care: A population-based cohort study | PLOS Medicine." <https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1003295> (accessed May 16, 2023).
- [5] "8775.00.pdf." Accessed: May 15, 2023. [Online]. Available: <https://www.cancer.org/content/dam/CRC/PDF/Public/8775.00.pdf>
- [6] E. Emmings, S. Mullany, Z. Chang, C. N. Landen, S. Linder, and M. Bazzaro, "Targeting Mitochondria for Treatment of Chemoresistant Ovarian Cancer," *Int. J. Mol. Sci.*, vol. 20, no. 1, p. 229, Jan. 2019, doi: 10.3390/ijms20010229.
- [7] A. Chandra et al., "Ovarian cancer: Current status and strategies for improving therapeutic outcomes," *Cancer Med.*, vol. 8, no. 16, pp. 7018–7031, 2019, doi: 10.1002/cam4.2560.
- [8] "Sci-Hub | Breast cancer diagnosis using multiple activation deep neural network | 10.1177/1063293X211025105." <https://sci-hub.wf/10.1177/1063293X211025105> (accessed May 17, 2023).
- [9] J. V. Tembhurne, A. Hazarika, and T. Diwan, "BrC-MCDLM: breast Cancer detection using Multi-Channel deep learning model," *Multimed. Tools Appl.*, vol. 80, no. 21–23, pp. 31647–31670, Sep. 2021, doi: 10.1007/s11042-021-11199-y.
- [10] S. K. Rajeev, M. Pallikonda Rajasekaran, G. Vishnuvarthanan, and T. Arunprasad, "A biologically-inspired hybrid deep learning approach for brain tumor classification from magnetic resonance imaging using improved gabor wavelet transform and Elmann-BiLSTM network," *Biomed. Signal Process. Control*, vol. 78, p. 103949, Sep. 2022, doi: 10.1016/j.bspc.2022.103949.
- [11] S. Srivastava, P. Kumar, V. Chaudhry, and A. Singh, "Detection of Ovarian Cyst in Ultrasound Images Using Fine-Tuned VGG-16 Deep Learning Network," *SN Comput. Sci.*, vol. 1, no. 2, p. 81, Mar. 2020, doi: 10.1007/s42979-020-0109-6.
- [12] X. Meng, X. Li, and X. Wang, "A Computationally Virtual Histological Staining Method to Ovarian Cancer Tissue by Deep Generative Adversarial Networks," *Comput. Math. Methods Med.*, vol. 2021, pp. 1–12, Jul. 2021, doi: 10.1155/2021/4244157.
- [13] Z. Fan et al., "Using machine learning to predict ovarian cancer," *Int. J. Med. Inf.*, vol. 141, p. 104195, Sep. 2020, doi: 10.1016/j.ijmedinf.2020.104195.
- [14] D. Schwartz, T. W. Sawyer, N. Thurston, J. Barton, and G. Ditzler, "Ovarian cancer detection using optical coherence tomography and convolutional neural networks," *Neural Comput. Appl.*, vol. 34, no. 11, pp. 8977–8987, Jun. 2022, doi: 10.1007/s00521-022-06920-3.
- [15] J. Yang, C. Xiang, and J. Liu, "Clinical significance of combining salivary mRNAs and carcinoembryonic antigen for ovarian cancer detection," *Scand. J. Clin. Lab. Invest.*, vol. 81, no. 1, pp. 39–45, Feb. 2021, doi: 10.1080/00365513.2020.1852478.
- [16] Z. Zhang and Y. Han, "Detection of Ovarian Tumors in Obstetric Ultrasound Imaging Using Logistic Regression Classifier With an Advanced Machine Learning Approach," *IEEE Access*, vol. 8, pp. 44999–45008, 2020, doi: 10.1109/ACCESS.2020.2977962.
- [17] J. D. Kaggie et al., "Feasibility of Quantitative Magnetic Resonance Fingerprinting in Ovarian Tumors for T 1 and T 2 Mapping in a PET/MR Setting," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 3, no. 4, pp. 509–515, Jul. 2019, doi: 10.1109/TRPMS.2019.2905366.
- [18] S. Negi, P. Mittal, and B. Kumar, "Modeling and Analysis of High-Performance Triple Hole Block Layer Organic LED Based Light Sensor for Detection of Ovarian Cancer," *IEEE Trans. Circuits Syst. Regul. Pap.*, vol. 68, no. 8, pp. 3254–3264, Aug. 2021, doi: 10.1109/TCSI.2021.3078510.
- [19] K. Kasture, "OvarianCancer&Subtypes." Mendeley, Mar. 31, 2021, doi: 10.17632/W39ZGKSP6N.1.
- [20] "Automatic Detection and Segmentation of Ovarian Cancer Using a Multitask Model in Pelvic CT Images." <https://www.hindawi.com/journals/omcl/2022/6009107/> (accessed May 18, 2023).
- [21] X. Wang et al., "Intelligent Hybrid Deep Learning Model for Breast Cancer Detection," *Electronics*, vol. 11, no. 17, Art. no. 17, Jan. 2022, doi: 10.3390/electronics11172767.
- [22] K. Srilatha and V. Ulagamuthalvi, "Support Vector Machine And Particle Swarm Optimization Based Classification Of Ovarian Tumour," *Biosci. Biotechnol. Res. Commun.*, vol. 12, no. 3, pp. 714–719, Sep. 2019, doi: 10.21786/bbrc/12.3/24.
- [23] N. Dhanachandra and Y. J. Chanu, "An image segmentation approach based on fuzzy c-means and dynamic particle swarm optimization algorithm," *Multimed. Tools Appl.*, vol. 79, no. 25–26, pp. 18839–18858, Jul. 2020, doi: 10.1007/s11042-020-08699-8.
- [24] C.-L. Huang, M.-J. Lian, Y.-H. Wu, W.-M. Chen, and W.-T. Chiu, "Identification of Human Ovarian Adenocarcinoma Cells with Cisplatin-Resistance by Feature Extraction of Gray Level Co-Occurrence Matrix Using Optical Images," *Diagnostics*, vol. 10, no. 6, Art. no. 6, Jun. 2020, doi: 10.3390/diagnostics10060389.
- [25] L. K. Kumari and B. N. Jagadesh, "A Robust Feature Extraction Technique for Breast Cancer Detection using Digital Mammograms based on Advanced GLCM Approach," *EAI Endorsed Trans. Pervasive Health Technol.*, vol. 8, no. 30, pp. e3–e3, Jan. 2022, doi: 10.4108/eai.11-1-2022.172813.
- [26] Q. M. Alzubi, M. Anbar, Z. N. M. Alqattan, M. A. Al-Betar, and R. Abdullah, "Intrusion detection system based on a modified binary grey wolf optimisation," *Neural Comput. Appl.*, vol. 32, no. 10, pp. 6125–6137, May 2020, doi: 10.1007/s00521-019-04103-1.
- [27] V. A. Chinnasamy and D. R. Shashikumar, "Breast cancer detection in mammogram image with segmentation of tumour region," *Int. J. Med. Eng. Inform.*, vol. 12, no. 1, pp. 77–94, Jan. 2020, doi: 10.1504/IJMEI.2020.105658.
- [28] R. Yazdani, M. Mirmozaffari, E. Shadkam, and M. Taleghani, "Minimizing total absolute deviation of job completion times on a single machine with maintenance activities using a Lion Optimization Algorithm," *Sustain. Oper. Comput.*, vol. 3, pp. 10–16, Jan. 2022, doi: 10.1016/j.susoc.2021.08.003.
- [29] V. A. Chinnasamy and D. R. Shashikumar, "Breast cancer detection in mammogram image with segmentation of tumour region," *Int. J. Med. Eng. Inform.*, vol. 12, no. 1, pp. 77–94, 2020.
- [30] P. H. Nagarajan and N. Tajunisha, "Optimal Parameter Selection-based Deep Semi-Supervised Generative Learning and CNN for Ovarian Cancer Classification," *ICTACT J. SOFT Comput.*, vol. 13, no. 02, 2023.
- [31] H. Su, E. Zio, J. Zhang, M. Xu, X. Li, and Z. Zhang, "A hybrid hourly natural gas demand forecasting method based on the integration of wavelet transform and enhanced Deep-RNN model," *Energy*, vol. 178, pp. 585–597, Jul. 2019, doi: 10.1016/j.energy.2019.04.167.
- [32] S. Bhanumathi and S. N. Chandrashekara, "DEEP EARNING BASED BiLSTM ARCHITECTURE FOR LUNG CANCER CLASSIFICATION," *Int. J. Adv. Res. Eng. Technol. IJARET*, vol. 12, no. 1, pp. 492–503, 2021.
- [33] B. Jang, M. Kim, G. Harerimana, S. Kang, and J. W. Kim, "Bi-LSTM Model to Increase Accuracy in Text Classification: Combining

- Word2vec CNN and Attention Mechanism,” *Appl. Sci.*, vol. 10, no. 17, p. 5841, Aug. 2020, doi: 10.3390/app10175841.
- [34] Z. Zhang and Y. Han, “Detection of ovarian tumors in obstetric ultrasound imaging using logistic regression classifier with an advanced machine learning approach,” *IEEE Access*, vol. 8, pp. 44999–45008, 2020.
- [35] G. Wadhwa, “A Deep Convolutional Neural Network Approach for Detecting Malignancy of Ovarian Cancer Using Densenet Model,” vol. 25, no. 2, 2021.
- [36] K. R. Kasture and E. Al, “A New Deep Learning method for Automatic Ovarian Cancer Prediction & Subtype classification,” *Turk. J. Comput. Math. Educ. TURCOMAT*, vol. 12, no. 12, Art. no. 12, May 2021.

# Utilizing Deep Convolutional Neural Networks and Non-Negative Matrix Factorization for Multi-Modal Image Fusion

Dr. Nripendra Narayan Das<sup>1</sup>, Santhakumar Govindasamy<sup>2</sup>,  
Dr. Sanjiv Rao Godla<sup>3</sup>, Prof. Ts. Dr. Yousef A. Baker El-Ebiary<sup>4</sup>, Dr. E. Thenmozhi<sup>5</sup>

Department of Information Technology, Manipal University Jaipur, Rajasthan, India<sup>1</sup>

Assistant Professor, Electronics and Communication Engineering, Sri Krishna College of Technology, Coimbatore, India-641042<sup>2</sup>

Professor, Department of CSE (Artificial Intelligence & Machine Learning),

Aditya College of Engineering and Technology-Surapalem, Andhra Pradesh, India<sup>3</sup>

Professor, Faculty of Informatics and Computing, UniSZA University, Malaysia<sup>4</sup>

Associate Professor, Department of Information Technology, Panimalar Engineering College, Chennai, India<sup>5</sup>

**Abstract**—A key element of contemporary computer vision, image fusion tries to improve the quality and interpretability of images by combining complimentary data from several image sources or modalities. This paper offers a unique method for multi-modal image fusion, combining the benefits of Deep Convolutional Neural Networks (CNNs) and Non-Negative Matrix Factorization (NMF), by using current developments in deep learning and matrix factorization techniques. Deep CNNs have shown to be remarkably effective in extracting features from images, capturing complex patterns and discriminative data. A group of deep CNNs are trained using this suggested technique on a varied dataset of multi-modal images. With the help of these networks, which extract and encode pertinent characteristics from several modalities, information-rich representations may then be combined. Concatenating, the features that were derived from the CNNs throughout the fusion process results in a fused feature representation that perfectly expresses the input modalities. The main novelty is the two-stage integration of NMF: first, breaking down the fused feature representation into non-negative basis vectors and coefficients, and then, using NMF to further extract important patterns from the fused feature maps. The non-negativity requirement in NMF guarantees the preservation of the natural structures and characteristics present in the source images, resulting in fused images that are both aesthetically pleasing and semantically intelligible. Visual examination of the merged images demonstrates the method's capacity to successfully extract important information from several modalities. The better performance and robustness of the suggested approach, which has an accuracy of roughly 99.12%, are highlighted by comparison with existing fusion approaches.

**Keywords**—Image fusion; deep convolution network; non-negative matrix factorization; multi-modal images; vector space model

## I. INTRODUCTION

Image fusion has a wide range of uses in both commercial and non-industrial sectors, including security. Due to technical or optical imaging limitations, only a portion of the information may be recorded in an image using a certain type of detector or firing configuration. For example, reflecting

illumination data that has intensity in a constrained range and falls within a predetermined depth-of-field is a classic example of insufficient data. By combining complimentary data gathered from many source images that were taken with various sensors or optical settings, image fusion aims to create a synthesized image [1]. Following visual assignments, including video monitoring, scene comprehension, target acknowledgment, etc., benefit from a single fusion image with greater environment representations and better perception of sight. It is challenging to efficiently and rapidly explore image on image-sharing systems due to the enormous volume of images submitted to services like Flickr and Picasa. This issue can be resolved using gathering images summarization. The goal of the image collection overview is to portray a huge, multi-modal library using only a small subset of the images and labels. The small subset shows the different elements of the initial collection, such as the attribute of interest and scene category. Image collection summarization may be employed for a variety of multimedia projects, such as automatic album building, search outcome optimization, etc. [2].

Different kinds of medical images serve an essential part in clinical diagnosis in contemporary medicine and are quite helpful in identifying disorders. Doctors typically need to integrate numerous different kinds of medical images from the same location in order to gather sufficient data for an appropriate evaluation, which frequently causes significant difficulty. When a doctor simply uses his or her own theories and conceptions to analyze a variety of medical visuals, the evaluation's objectivity is compromised and it's possible that some of the image's data is overlooked. Techniques for image fusion offer a practical solution to these problems. The collected healthcare images from various modes contain supplementary as well as duplication of data as the range of medical imaging technologies grows [3]. Other research has used a combination of verbal and graphic data to create image representations [4]. To create the visual short, the investigator developed an overview challenge that involved locating subset image examples using a homogeneous and heterogeneous message transmission technique. The image summary challenge was transformed into a hyper-graph division issue

through research, which took into account both visual and textual aspects. It suggested a max-margin assistance vector machine-based technique to extract visual ideas from multimodal dataset [5].

An imaging equipment, such as a digital single-lens reflex the device, frequently finds it challenging in the field of electronic imaging to take an image in which all the objects are sharply focused [6]. Only subjects in the depth-of-field (DOF) of an optically lens will usually look crisp in a shot at a given focal length; subjects outside the DOF will most likely to be blurry. Multi-focus image fusion, which combines many images of the same subject captured at various focal lengths to create an all-in-focus appearance, is a common approach. Additionally, a significant area within the field of image fusion is multi-focus fusion of images [7]. Many techniques for merging multi-focus images may be used, regardless of alterations, for additional image fusion applications like visible-infrared image fusion and multi-modal healthcare image synthesis. Investigating multi-focus image fusion has dual significance from this perspective, making it an explosive subject in the image computing field. Several image fusion techniques have been developed in recent years, and these techniques may be loosely divided into two distinct groups: transformation area techniques and spatially domain technique [8]. Data fusion and mining include integrating and analysing many data sources to draw out insightful conclusions and patterns. To get a complete picture of the data landscape, it aggregates data from multiple heterogeneous sources, including databases, sensors and social media. When data from many sources are combined, conflicts are resolved, and a single dataset is produced that more accurately depicts the underlying phenomenon [9]. Then, using data mining techniques, relevant relationships, trends, and patterns are identified from the pooled data. Using techniques from machine learning as well as statistical methodologies, anomalies, hidden trends, and predictions or suggestions are found in this process. Data fusion and mining are widely used in a variety of industries, such as health care, banking, marketing, and cybersecurity, and they allow companies to make data-driven decisions that improve productivity, efficiency, and decision-making [10].

To generate the multi-modal overview successively, Camargo and Gonzalez i [11] used convex non-negative matrix factorization (convex NMF) to visual modalities expressed as BoW and textual modalities expressed as vector space model (VSM). However, they did consider the sequential association between the images and labels. The characteristics of the literary topic were first taken into account. Next, images were used as inspiration for the visual concept. The sequential method, however, limits the dissemination of information from various data. They primarily relied on the textual aspects of visual summaries, ignoring the visual aspects of the literary issue and the diverse interactions between the two mediums. As a result, older summary techniques are unable to generate outcomes that exactly match the initial collection. Early spatial domain approaches frequently employed block-based fusion. Depending on the subject of the images, blocks of various sizes can be created adaptively from the images. The concept of block-based techniques is shared by a different

class of spatial domain techniques that rely on image segmentation. However, the effectiveness of the classification has a big influence on how effectively these tactics work together. Many unique gradient-based, pixel-based spatial domain approaches have been created recently that can produce state-of-the-art multi-focus image fusion results. These approaches usually use rather complex fusion strategies (which can be thought of as rules in a broad sense) to their computation findings from activity degree analysis in order to boost the fusion efficacy.

The key contributions of the Multi-Modal Image Fusion using Deep Convolutional Neural Networks (CNNs) and Non-Negative Matrix Factorization (NMF) approach are:

- By concatenating the features obtained from the CNNs during the fusion process, a fused feature representation is produced that accurately captures the essence of the input modalities, improving image quality and interpretability.
- Deep CNNs are used to extract features from multi-modal images, showcasing their exceptional ability to capture intricate patterns and discriminative data, which are crucial for producing informative fused images.
- The integration of NMF in two stages is the primary innovation. In order to improve the fusion process, two steps must be taken: first, the fused feature representation must be broken down into non-negative basis vectors and coefficients; and second, NMF must be utilized to extract important patterns from the fused feature maps.
- The non-negativity condition in NMF makes sure that the fused images retain the organic shapes and traits that are present in the source images, resulting in fused images that are both aesthetically pleasing and semantically significant.

This article's remainder is organized as follows: In Section II, a summary of related research is provided. Section III presents the problem statement. The suggested approach's methodology and architecture are explained in Section IV of the article. The findings and subsequent discussion are covered in Section V. The conclusion is covered in Section VI.

## II. RELATED WORK

Although multi-model neuroimaging and gene identification technologies have advanced, attempts to integrate the two in order to investigate the virulence traits of schizophrenia (SZ) have been unsuccessful. Researchers suggest a unique approach known as grouping dense of joint non-negative matrices factorization on orthogonal domain to address this problem. The approach combines data from three models, single nucleotide polymorphism, and functional magnetic resonance imaging to identify risk genes, aberrant brain areas, and SZ-related epigenetic elements. For the purpose of eliminating unnecessary characteristics from the row of correlation matrix structures, researchers actively place diagonal constraints on the foundation matrix. Because data from genome-scanning provide extensive group information, researchers use three coefficients vectors that are densely

packed to enhance the features discovered. Our approach is tested using both the made-up and actual Mind Clinical Imaging Consortium (MCIC) datasets. Simulation results demonstrate that our approach outperforms rival tactics. GJNMFO identifies a set of risk genes, epigenetic variables, and aberrant brain functioning regions through the use of MCIC data in the study. These findings have significant economic and ecological ramifications, which science has proven [12].

For ground-based cloud recognition, deep neural networks have recently attracted a lot of attention. The entire focus of these techniques, however, is on extrapolating global features from visual input, which results in approximations for ground structures that are erroneous. The multi-evidence and multi-modal fusion network (MMFN), which is described in this article, is a unique technique for ground-based cloud identification that can increase cloud knowledge by fusing various signals in an integrated framework. By utilizing the attentive network and the main system, MMFN specifically uses a number of data points, such as global and local visual characteristics, from ground-based clouds images. Local visual qualities are gathered using the attention maps of the attentive system, which are constructed using fine-tuned salient aspects of convolutional stimulating structures. The multi-modal networking in MMFN is now studying the multi-modal properties of ground-based clouds. Researchers developed two fusion stages in MMFN to combine multi-modal features with local and global visual properties in order to fully fuse the multi-modal and multi-evidence visual qualities. The first multi-modal ground-based cloud database, or MGCD, is also made possible by study. It includes both the ground-based cloud images themselves as well as the multi-modal data that goes with each cloud image. When measured against state-of-the-art techniques, the MMFN obtains an identification performance of 88.63% on MGCD, proving its suitability for ground-based cloud recognition. The current study forbids the use extra factors, such as cloud basal height, for cloud characterization [13].

To achieve human-robot collaboration (HRC) in manufacturing processes, multimodal robot management must be intuitive and trustworthy. In earlier works, multimodal robotic control strategies were established. The technologies make it possible for human employees to control robots naturally without having to write brand-specific programming. However, because characteristics are not depicted consistently across multiple paradigms, the bulk of multimodal controlling robots' approaches are unreliable. In order to solve this problem, the research on reliable multimodal HRC production systems suggests a multimodal fusion architecture that makes use of deep learning. The proposed design consists of three modalities: verbal authority, hand gesture, and body motion. Three single-modal systems' characteristics are first trained to be retrieved, after which the characteristics are combined to swap representations. Tests show that the proposed multimodal fusion paradigm performs superior to the three unimodal models. The paper emphasizes the potential for applying the suggested multimodal fusion architecture to produce dependable HRC systems. The architectural concept paradigm wasn't made clear enough [14].

Radar electronic surveillance has new difficulties as a result of the emergence of cognitive wireless and electronic warfare; recognizing the signal generated by radar is a crucial component of this work. Research suggests a new radar signal recognition technique that uses non-negative matrix factorization network (NMFN) and ensemble learning. This system is capable of reliably recognizing radar signals under low signal-to-noise ratio conditions. Research investigates a method for extracting features based on a convolutional neural network at the beginning, which uses transfer learning as a way to address the issue of small sizes of samples. In order to extract characteristics and eliminate redundant data, research also suggests a non-negative matrix factorization system. In the third step, research create a feature fusion method utilizing stacked autoencoders (SAE), which can collect key feature expressions and condense feature dimensions. Last but not least, researcher suggests the improved artificial bee colony algorithm (IABC) as an ensemble learning technique that can increase the recognition rate. According to the simulation outcomes, recognition rates are 94.23% at 4 dB and 99.82% at 6 dB [15].

Dynamic MRI was used as a technique to record the body's various organs successive anatomy as they change over time. Nevertheless, due to mechanical and physiological limitations, its uses are restricted by shorter acquisition times. It has been demonstrated that dynamic MRI has spatio-temporal heterogeneity in its frequency spectrum (k-space). Lowering the number of k-space examples can greatly shorten the acquisition duration, yet at the expense of introducing artefacts into the associated image realm. To speed up the whole acquisition procedure, Shashidhar and Subha [16] created a cascaded Convolutional Long Short-Term Memory (ConvLSTM) framework for T2-weighted dynamic MRI patterns restoration from significantly under-sampled k-space information. Particularly, a Cartesian inadequate sampling mask could be used to under-sample completely sampled information obtained from the ADNI dataset. The aliasing artefacts caused by inadequate sampling are then eliminated using the ConvLSTM framework that has been suggested. In order to rebuild the imagery effectively and more accurately than CNN-based restoration, the ConvLSTM framework also learns the imagery's temporal and spatial connections. The utilization of medical databases presents ethical issues about data protection and informed approval, like the ADNI database. It is crucial to confirm that the research complies with ethical standards and has gotten the necessary rights and authorization for the use of the data.

### III. PROBLEM STATEMENT

The requirement for efficient multi-modal image fusion to improve image quality and interpretability across many applications is the issue this research attempts to solve. Integrating data from several sources while preserving the accuracy of the original data are difficult. Complex patterns and distinguishing traits are frequently difficult to capture using traditional techniques. The work suggests a remedy that combines the capacities of CNNs and NMF to address this. Utilizing CNNs' feature extraction abilities, the idea is to produce a fused feature representation that is then improved by NMF to uncover useful patterns. Since the non-negativity

requirement in NMF, the fused images are coherent and meaningful since their inherent structures are preserved. The suggested approach's capacity to maintain crucial diagnostic data and improve fusion quality is assessed by quantitative metrics and visual evaluations, demonstrating its potential to help precise decision-making and analysis in areas like medical imaging [17].

#### IV. PROPOSED FRAMEWORK

There are multiple steps in the suggested methodology for Deep Convolutional Neural Networks (CNNs) to fuse multimodal images. The process for Multi-Modal Image Fusion using Deep Convolutional Neural Networks and Non-Negative Matrix Factorization is depicted in Fig. 1. The input images 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 images and a fusion-specific loss function, 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. NMF is used to decompose extracted feature maps into non-negative basis vectors and coefficients in the setting of non-negative matrix factorization (NMF) feature extraction. The most discriminative basis vectors capturing the essential features of each modality are selected, and fusion weights are learned or fusion rules are applied to combine them. The fused basis vectors are multiplied with the corresponding coefficients to reconstruct the fused feature maps, which are then aggregated to generate the final fused image. Post-processing techniques like denoising or sharpening can be applied for further enhancement. This methodology is primarily used for multi-modal image fusion, where images from different modalities, such as infrared and visible, are combined to provide a comprehensive understanding of a scene. On the other hand, non-multi-modal image fusion is utilized in feature fusion scenarios where features from the same modality but captured under different conditions, such as exposure or focus, are fused to create a more comprehensive feature representation.

##### A. Data Collection

MRI brain images of 1000 datasets including healthy and unhealthy are used in the research. Among these 50% of images are used as training data and 50% of images are used as testing data. The collected brain cancer images were existing on the Kaggle depository website [18]. The datasets are distributed in Table I.

##### B. Data Preprocessing

Magnetic Resonance Imaging (MRI) imageries were impacted by unrelated and erratic noisy data, such as Gaussian noise and Speckle sound, which reduced the analysis value of those sample imageries. Speckle sounds have a significant impact on the contrast resolution of MRI brain imaging. Therefore, the original Hannmean filter is used to reduce noise in MRI brain images. A Hannmean filter is a filter that combines the Hanning window and Mean filters. The established Hannmean filter is used to minimize the noise in an image as well as any spatial intensity derivatives that may be present. In order to exchange each pixel's value with its surrounding neighbours' mean image values and ignore the

unreliable pixel value of their image background, the Hannmean filter is used. Noises are produced in MRI brain scans by the device's inhomogeneity dis a magnetic region afforded by temperature, the malfunction of the scanner, and the patient's movement throughout the scanning process. Both noiseless methods and image resolution were used to get a crisp MRI brain image [19].

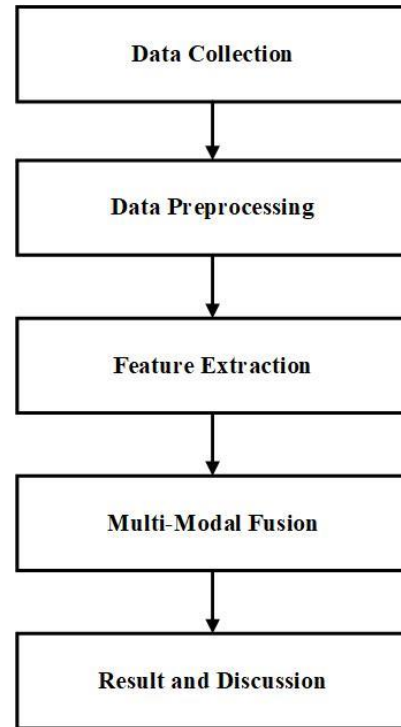


Fig. 1. Proposed framework.

TABLE I. THE COLLECTED DATASETS

	Training data	Testing data
Unhealthy	250	250
Healthy	250	250
Overall data	500	500

##### C. Feature Extraction using CNN

Deep CNNs are remarkably good at capturing hierarchical representations and complicated patterns. In order to begin the feature extraction process, a series of deep CNNs are trained on a variety of datasets made up of multi-modal images. These networks learn to recognize distinguishing elements that are pertinent to each input modality since they are designed to the specifics of the input modalities. CNNs extract features that capture detailed textures, forms, and structures unique to each modality by utilizing both low-level and high-level filters. The cross-modal linkages are preserved while modality-specific subtleties are captured in the learnt features. Concatenating the retrieved features from the different CNNs results in the formation of the fused feature representation, this completely embodies the essence of the multi-modal inputs. The basis for further processing, such as the usage of NMF to hone and extract more abstract patterns, is this fused feature representation. The suggested framework improves the fusion

process by utilizing CNNs' advantages in extracting valuable features, thereby assisting in the creation of high-quality fused images that capture the complimentary information available in the multi-modal data.

The suggested medical image fusion architecture contains three primary phases, which are depicted in Fig. 2. To begin, it creates the same-size weight map (m) for source images X and Y of arbitrary size using Siamese network architecture. The produced weight map X is then subjected to Gaussian gradient deconstruction to produce the matching multi-scale sub-decomposed image Gm, which is used to establish the fusion operation in the coefficient calculation merger procedure. The top layer and the remaining layers of the sub-decomposed

image are represented by the symbols  $G_{M,i=N}^{i,s}$  and  $G_{M,0\leq1\leq N}^{i,s}$  [20]. The contrast gradient is used to break down the source images X and Y. For the following coefficient fusion technique, the multi-scale sub-decomposed images  $Q_x$  and  $Q_y$  are acquired. The top layers of the sub-decomposed images  $Q_x$  and  $Q_y$  are  $Q_{X,i=N}^{i,s}$  and  $Q_{Y,i=N}^{i,s}$  correspondingly. In order to denote the other layers of the sub-decomposed images  $Q_x$  and  $Q_y$ , accordingly, research adopts the notation  $Q_{X,i=N}^{i,s}$  and  $Q_{Y,0\leq1\leq N}^{i,s}$ . Finally, distinct thresholds are established, one for the top level and the other for the layers that make up sub-decomposed image  $F_q$ .

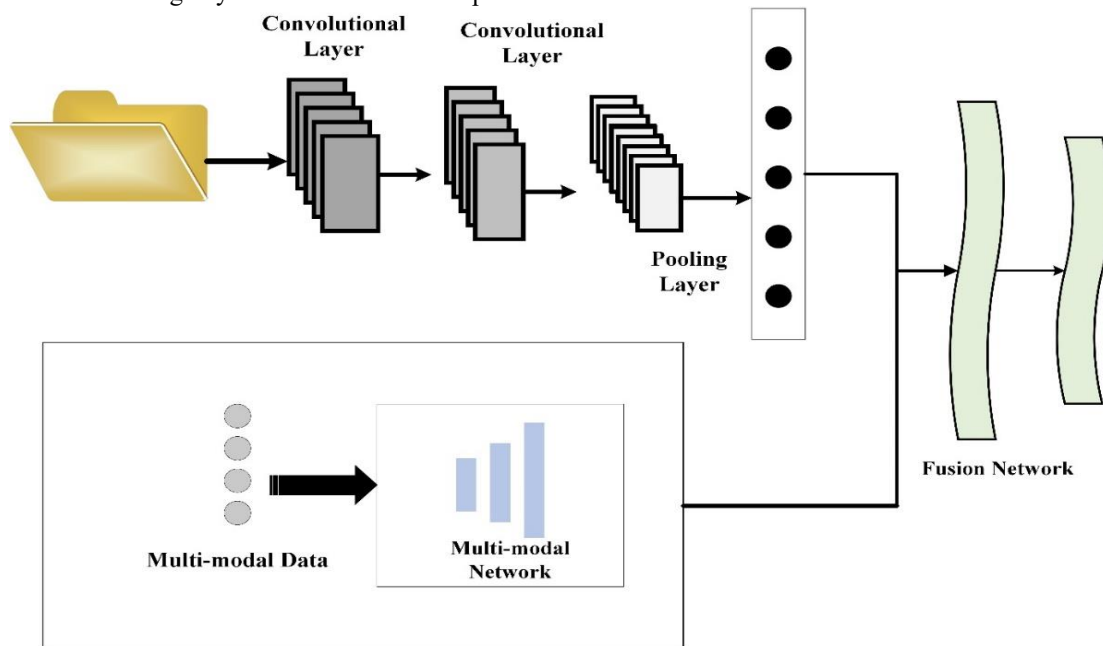


Fig. 2. Multi-modal fusion based deep convolution network.

The Fig. 2 represents a Multi-Modal Image Fusion based Deep Convolution Network, a powerful technique that combines information from different image modalities to generate a fused image with improved quality and interpretability. The network consists of input layers for each modality, followed by convolutional layers that extract relevant features from the images. Pooling layers down sample the feature maps, while fusion layers combines the extracted features from different modalities to create a comprehensive representation. Fully connected layers further transform and abstract the fused features, leading to an output layer that generates the final fused image. This architecture allows the network to leverage the strengths of each modality and enhance the understanding of the scene, making it a valuable tool in various applications [21].

The suggested technique uses CNN to accomplish an estimation of the ideal pixel level of activity and distributed weight by obtaining a weighted map of pixel activity details from numerous source images. In this study, Siamese networks are used to increase the effectiveness of CNN instruction. The Siamese system has two divisions. There are three convolutional layers and one max-pooling layer on each

branch. Convolutional neural networks comprise the top two layers. The input image's non-negative matrix factorization feature extraction is done on the first layer. There are more feature maps in the second layer. The top convolutional layer extracts the characteristics of the output map. In the proposed methodology, a max-pooling layer is included as the third layer in the network architecture. This layer serves to further reduce the number of parameters and remove unnecessary samples from the feature map. By down sampling the input feature map, the max-pooling layer retains the most significant information while discarding less relevant details, effectively reducing the computational complexity. Following the max-pooling layer, a fourth layer is introduced as a convolution layer. This layer extracts more intricate and detailed information from the pooled feature map, capturing finer patterns and features. To minimize the training complexity and memory usage, a lightweight network structure is employed for this convolutional layer [22].

Specifically, the feature maps from each branch are concatenated together in the network's final stage. Concatenation gives a more thorough representation by allowing the integration of knowledge gained from several

branches. The concatenated feature maps are then immediately coupled to a two-dimensional vector using a completely connected layer. The next bi-directional SoftMax layer uses this vector as its input. The two-dimensional vector is categorized based on probability values by the bi-directional SoftMax layer, which also forecasts the probability distribution of several qualities. This forecast is essential for estimating the probability that various attributes will appear in the input data. In order to predict attributes, the network outputs a probabilistic classification by mapping the two-dimensional vector to the SoftMax layer. Overall, to extract and express complicated information from the input data, this technology combines max-pooling, convolutional layers, and concatenation of feature maps. Accurate attribute prediction based on learned representations is made possible by the fully connected layer, the bi-directional SoftMax layer, and the probability-based categorization. This method is appropriate for a variety of applications requiring attribute prediction or classification since it minimizes the number of parameters, optimizes training complexity, and increases memory efficiency [23].

This study uses a SoftMax classifier to determine the categorization probability using Eq. (1) in order to achieve categorization in the DCNN network:

$$f(r_u) = \frac{e^{r_u}}{\sum_{v=1}^n e^{r_u}} \quad (1)$$

The mapping between each element of one ( $r_u$ ) will be approximately 1 and the rest will be closest to 0, normalizing all input matrices if one  $\pi_i$  is greater than every other component  $r$ . The SoftMax loss curve is found as Eq. (2) when the number of batches is set to 128:

$$B = \sum_{u=0}^{size} -\log f(r_u) \quad (2)$$

Stochastic gradient descent is utilized to minimize the loss function with the SoftMax loss value serving as the optimization objective. The acceleration loss and weight decay are established as the initial parameter values, respectively. Consequently, the weights are updated using Eq. (3):

$$s_{u+1} = s_u + t_u + 1 \quad (3)$$

In Eq. (3) the dynamic factor is defined as  $t_u$  and the weight is denoted as  $s_u$  at  $u^{\text{th}}$  iteration.

#### D. NMF for Multi Modal Image Fusion

By permitting the breakdown of fused feature representations into non-negative basis vectors and coefficients, NMF plays a crucial role in the field of Multi-Modal Image Fusion. This decomposition technique perfectly reflects the properties of image data, where pixel values are always positive. NMF makes it easier to create a fused image while maintaining the underlying natural structures and attributes existing in the input modalities by imposing this non-negativity requirement. The basis vectors, which depict fundamental patterns shared by all modalities, identify crucial characteristics that are similar to all inputs. By permitting the breakdown of fused feature representations into non-negative basis vectors and coefficients, NMF plays a crucial role in the field of Multi-Modal Image Fusion. This decomposition technique perfectly reflects the properties of image data,

where pixel values are always positive. NMF makes it easier to create a fused image while maintaining the underlying natural structures and attributes existing in the input modalities by imposing this non-negativity requirement. The basis vectors, which depict fundamental patterns shared by all modalities, identify crucial characteristics that are similar to all inputs.

Non-negative matrix factorization (NMF) is a powerful technique for multi modal fusion that aims to decompose a given data matrix into two non-negative matrices: a basis matrix and a coefficient matrix. In the context of feature extraction, NMF allows the extraction of meaningful and interpretable features by representing the input data as a linear combination of basis vectors. The basis matrix captures the fundamental components or patterns present in the data, while the coefficient matrix indicates the contribution of each basis vector to reconstruct the original data [24]. NMF assures that the extracted features are additive and non-competitive by applying non-negativity restrictions. This can be helpful for a variety of applications, including text mining, audio analysis, and image processing. The resulting basis vectors give the input data a condensed representation by emphasizing the key traits and bringing down the dimensionality, making it easier to perform further analysis or classification tasks. Overall, NMF-based feature extraction provides a practical method for identifying latent characteristics in data, enhancing the representation, comprehension, and use of complicated datasets [25].

For the assessment of non-negative matrices, the non-negative matrix factorization is used.  $A \in U_{G \times U}^+$  and  $B \in U_{U \times J}^+$  in which the two-matrix multiplication is similar to non-negative matrix  $C \in U_{G \times U}^+$  could be computed using the Eq. (4):

$$C = AB + F \quad (4)$$

Where  $F \in U_{G \times J}$  is an error matrix. The cost function connecting C and AB is minimized to predict the matrix of A and B as:

$$A = \arg \min_A Y(C|AB) \text{ for fixed } B \quad (5)$$

$$B = \arg \min_B Y(C|AB) \text{ for fixed } A \quad (6)$$

In Eqs. (5) and (6) the space between the two matrices of K and L is defined as  $V(K|L)$ .

The magnitude spectrogram of the signals is frequently used as the input matrix I in various applications of non-negative matrix factorization (NMF) for acoustic signals. In this instance, the frequency content of the acoustic wave over time is represented by the matrix I. Two non-negative matrices, A and B, are created by factorizing the matrix V. The spectrum features are represented by the matrix A, where each column vector represents a particular frequency structure or spectral component. The matrix B, on the other hand, reflects the temporal activations of acoustical events. Each row vector in this matrix represents the temporal envelope of a particular event. Research has been able to roughly reconstitute the magnitude spectrogram by multiplying matrices A and B. Consider a musical signal made up of three musical events to demonstrate this idea. Each column vector in matrix A would



be able to represent a different spectral pattern or frequency structure connected to the occurrences. The temporal envelopes of the various musical events would be represented by the row vectors of the matrix B, which would show how their amplitudes changed over time.

A conversion phase is used during the image testing and fusion process on the entirely linked layer to allow processing of sources of any size. The fully connected layer is converted into two identical convolutional layers with the same kernel size. Afterwards, the network may process any size images X and Y together in order to create a dense prediction map I. Each forecast  $I_s$  on the map has a two-dimensional vector with values ranging from 0 to 1. To make the weights assigned to corresponding image blocks simpler, if one dimension of a prediction is larger than the other, it is normalized to 1 while the other dimension is set to 0. This ensures that the weight of every image block is decreased with an output dimension value of 1. In their related image blocks, two close forecasts in S have overlapped areas. The mean value of the overlapping image blocks is obtained by adding the weights of the images in these overlapped sections. With the help of this method, the network can be fed images of any size, both X and Y, and a weight map W of the same size is produced. This makes sure that each image block's weight is reduced with an output dimension value of 1. The linked image blocks of the two close forecasts in I have overlap sections. The weights of the images in these overlapped portions are added to determine the mean value of the overlapping image blocks. Using this method, the system can produce a weight map W that is the same size as an image and accept images of any size, X and Y [19].

## V. RESULT AND DISCUSSION

Accuracy, Recall, Precision, F1-score, False Detection rate, Sensitivity, and Specificity are a few of the metrics used to verify the effectiveness of the projected model. True positive ( $t_p$ ), false negative ( $f_n$ ), false positive ( $f_p$ ), and true negative ( $t_n$ ) values are the fundamental variables that need to be computed.

### A. Accuracy

It gauges how precisely the system paradigm functions. In general, it refers to the ratio of correctly observed measurements to all data. The accuracy is presented in Eq. (7) as,

$$Accuracy = \frac{t_p + t_n}{t_p + t_n + f_p + f_n} \quad (7)$$

### B. Precision

The number of right positive estimates multiplied by the total number of positive guesses is used to measure precision. It is the percentage of precisely fused multi-modal medical images. Using Eq. (8), the precision is calculated as,

$$Precision = \frac{t_p}{t_p + f_p} \quad (8)$$

### C. Recall

Recall is defined as the ratio of true positives and false negatives to correct positive forecasts. It indicates the percentage of predictions that were accurate. multiple-modal image fusion. Eq. (9) is used to represent recall:

$$Recall = \frac{t_p}{t_p + f_n} \quad (9)$$

### D. Sensitivity

It is a measure of the proportion of correctly foretold true positives. Eq. (10) is used to calculate sensitivity as,

$$Sensitivity = \frac{t_p}{t_p + t_n} \quad (10)$$

### E. Specificity

The degree gauges how many precisely identifiable true negatives there are. Eq. (11) is used to calculate the specificity value as,

$$Specificity = \frac{t_n}{f_p + t_n} \quad (11)$$

TABLE II. COMPARISON OF PERFORMANCE METRICS

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Wavelet Transform	98.34	93.12	94.36	98.33
Fuzzy Logic	97.55	96.77	95.76	97.52
PCA	98.11	98.14	97.87	96.85
Proposed CNN-NMF	99.12	98.56	98.33	98.25

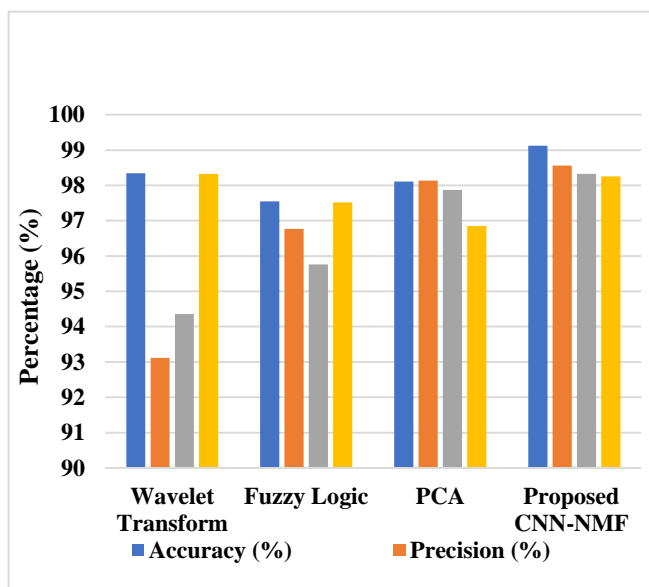


Fig. 3. Comparison of existing and proposed methods.

The Table II displays the accuracy, precision, recall, and F1-score performance evaluation of several image fusion techniques. The suggested CNN-NMF fusion strategy stands out among the tested techniques with the best accuracy of 99.12%, illustrating its capacity to successfully integrate multi-modal data. Additionally, this approach achieves impressive accuracy, recall, and F1-score values of 98.56%, 98.33%, and 98.25%, demonstrating its competence in accurately recognizing positive cases and minimizing false positives and negatives. The proposed CNN-NMF approach outperforms competing techniques like Wavelet Transform, Fuzzy Logic, and PCA, but it also has the potential to improve multi-modal image fusion tasks by capturing intricate patterns and preserving the integrity of the original data. It is depicted in Fig. 3.

TABLE III. MEDICAL IMAGE FUSION COMPARISON

Methods	Tsallis entropy	Gradient-based quality	Information ratio	Mutual information	Processing Time
MST-SR	64%	39%	36%	97%	15.05
NSCT-PC	71%	44%	40%	90%	3.77
ASR	66%	35%	39%	68%	6.15
CNN-LIU	62%	62%	28%	87%	14.58
Proposed	98%	45%	41%	92%	12.86

For the fused model with trainable and non-trainable weights, Fig. 4 and 5 displays the training accuracy and loss. Fig. 4 and 5 can be compared, and it is obvious that the model with trainable weights exhibits a faster improvement in accuracy and loss than the model with non-trainable weights. However, both networks achieve a point of convergence after around 40 epochs, with a training accuracy of about 98.07% and a loss of 0.0496. The fused model achieves a remarkable accuracy of 99.58% for the test dataset. These results show that both models eventually perform at a similar level in terms of accuracy and loss, however the model with trainable weights shows faster early development

A comparison of various methodologies based on various evaluation indicators and processing time is presented in Table III and Fig. 6. The NSCT-PC, CNN-LIU, ASR, MST-SR, and proposed algorithms are the ones that were tested. The Proposed technique receives the best score of 98% for Tsallis entropy, demonstrating its efficacy in maintaining information during the fusion process. While MST-SR, ASR, and CNN-LIU score lower with 64%, 66%, and 62% correspondingly, NSCT-PC comes in second with 71%. CNN-LIU receives the greatest score for gradient-based quality (62%), demonstrating its capacity to catch fine gradients in the fused image. The Proposed technique and ASR score 45% and 35%, respectively, whereas NSCT-PC scores 44%. The Proposed method achieves a 41% information ratio, showing a balanced preservation and utilisation of information. Following with 40% is NSCT-PC, followed by ASR with 39% and CNN-LIU with 28%. The Proposed technique receives a 92% for mutual information, demonstrating a high degree of mutual dependence between the input images in the fused result.

Following with 90% is NSCT-PC, and CNN-LIU comes in at 87%. The scores for MST-SR and ASR are lower, at 97% and 68%, respectively. In terms of processing speed, NSCT-PC performs the best with a time of 3.77. The Proposed approach achieves 12.86, whereas ASR comes in second with 6.15. The processing times for MST-SR and CNN-LIU are 15.05 and 14.58, respectively. The Proposed method achieves competitive scores for gradient-based quality and stands out in terms of Tsallis entropy, information ratio, and mutual information. A promising method for multi-modal image fusion, it surpasses competing techniques in most assessment measures despite taking a little longer to process data than NSCT-PC.

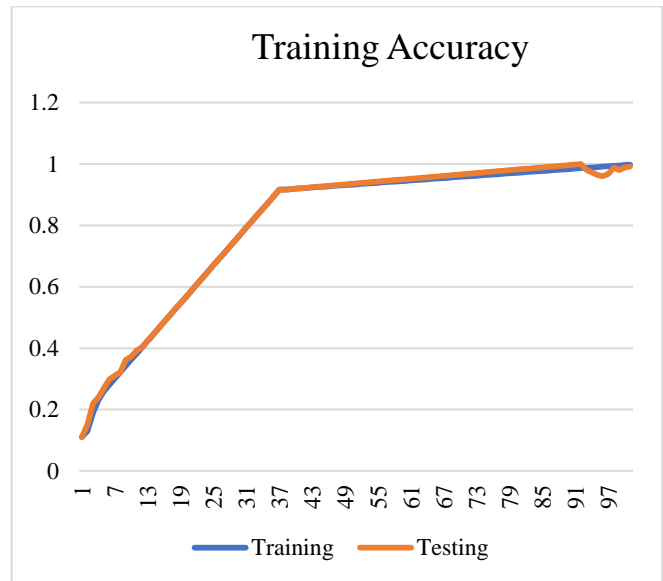


Fig. 4. Training accuracy.

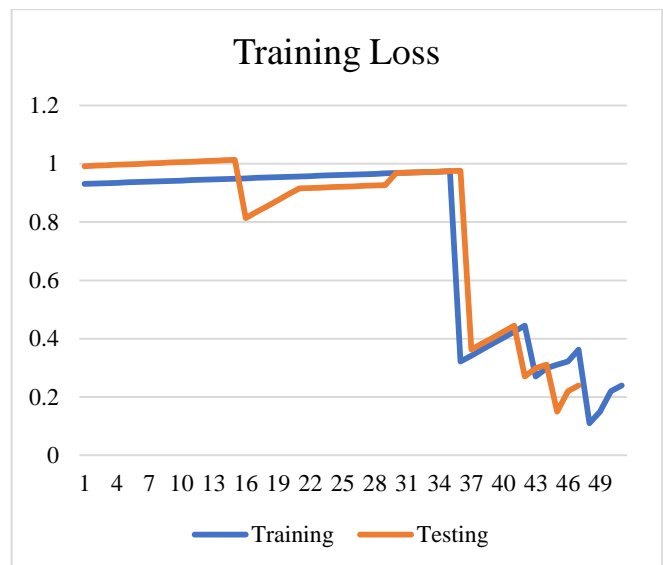


Fig. 5. Training loss.

F. Discussion

The proposed research offers a novel method for multi-modal image fusion within the context of modern computer

vision that makes use of both Deep CNNs and NMF's advantages. Deep CNNs have been shown to be adept at extracting minute details and identifying subtle patterns in images, making them an invaluable tool for working with multi-modal data. The paper suggests using these deep CNNs in a two-stage fusion process. First, the neural networks are trained to extract significant features from various modalities, and then the features that were extracted are concatenated to provide a thorough fused picture of the input data. This method stands out due to the creative ways in which NMF is applied in two different phases: first, to break down the fused representations of features into non-negative basic vector and coefficients, and subsequently, to further extract significant patterns from the resulting fused feature maps. The inherent non-negativity requirement in NMF guarantees the preservation of organic structures and inherent qualities in the source images, producing fused images that are visually beautiful and semantically comprehensible. The method excels at extracting critical information from many modalities, as shown by a visual analysis of the fused images. Its amazing accuracy also stands out as a noteworthy accomplishment, beating other fusion techniques and demonstrating its better performance and resilience. As a result of the partnership between deep CNNs and NMF, this work offers an appealing method for multi-modal picture fusion that yields a reliable and highly precise fusion technique. The suggested method successfully collects and combines data from several modalities, producing combined images that are not only aesthetically pleasing but also semantically relevant. This is made possible by successfully merging both cutting-edge deep learning methods with matrix factorization techniques. This multi-modal fusion invention is poised to make major strides in a number of sectors that depend on picture processing and interpretation and where it is crucial to accurately extract complementing information from many sources.

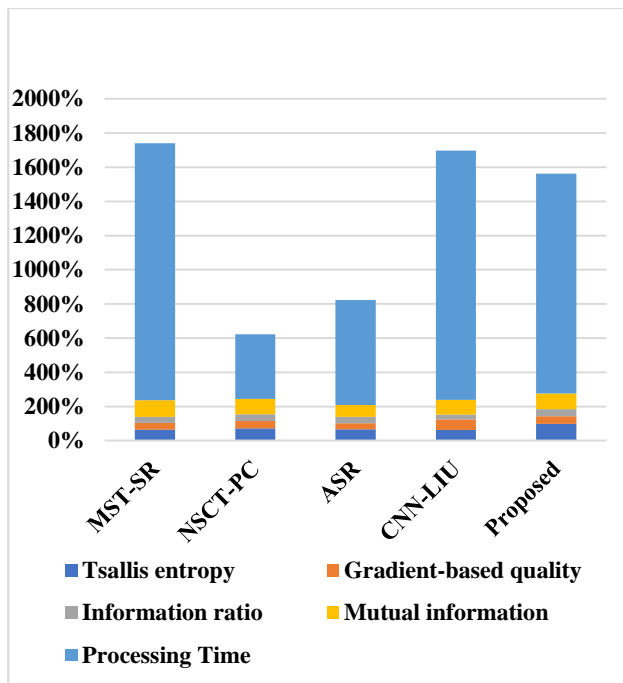


Fig. 6. Objective evaluation comparison.

## VI. CONCLUSION

In this research, a unique and efficient multi-modal image fusion approach that makes use of Deep CNNs and NMF is provided. The suggested method tackles the fundamental problem of improving image quality and interpretability through fusion by taking use of current developments in deep learning and matrix factorization techniques. Deep CNNs have been shown to be effective in extracting features from a variety of input modalities, underscoring its importance in this situation by capturing complex patterns and discriminative data necessary for successful fusion. The approach creates information-rich representations that are then smoothly merged via the fusion process by training a series of deep CNNs on a variety of datasets. By allowing the extraction of crucial patterns from fused feature representations while conserving the inherent structures of the source images, the dual-stage integration of NMF represents a singular invention. This preservation, which is grounded in NMF's non-negativity condition, produces fused images that are both aesthetically cohesive and semantically understandable. The visual proof of information effectively collected from many modalities supports the approach's potential even more. This study's overall findings represent a substantial improvement in multi-modal image fusion, with potential applications in industries that need precise data integration and image enhancement.

## VII. REFERENCES

- [1] J. Gao, Y. Lu, J. Qi, and L. Shen, "A radar signal recognition system based on non-negative matrix factorization network and improved artificial bee colony algorithm," *IEEE Access*, vol. 7, pp. 117612–117626, 2019.
- [2] F. Behrad and M. Saniee Abadeh, "An overview of deep learning methods for multimodal medical data mining," *Expert Syst. Appl.*, vol. 200, p. 117006, Aug. 2022, doi: 10.1016/j.eswa.2022.117006.
- [3] S. Zheng, B. Fang, L. Li, M. Gao, Y. Wang, and K. Peng, "Automatic liver lesion segmentation in CT combining fully convolutional networks and non-negative matrix factorization," in *Imaging for Patient-Customized Simulations and Systems for Point-of-Care Ultrasound: International Workshops, BIVPCS 2017 and POCUS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, 2017, Proceedings, Springer, 2017*, pp. 44–51.
- [4] K. C. Ravikumar, P. Chiranjeevi, N. Manikanda Devarajan, C. Kaur, and A. I. Taloba, "Challenges in internet of things towards the security using deep learning techniques," *Meas. Sens.*, vol. 24, p. 100473, Dec. 2022, doi: 10.1016/j.measen.2022.100473.
- [5] A. Khalil, M. Elmogy, M. Ghazal, C. Burns, and A. El-Baz, "Chronic wound healing assessment system based on different features modalities and non-negative matrix factorization (nmf) feature reduction," *IEEE Access*, vol. 7, pp. 80110–80121, 2019.
- [6] B. Lin, X. Tao, and J. Lu, "Hyperspectral image denoising via matrix factorization and deep prior regularization," *IEEE Trans. Image Process.*, vol. 29, pp. 565–578, 2019.
- [7] B. Swiderski, J. Kurek, S. Osowski, M. Kruk, and W. Barhoumi, "Deep learning and non-negative matrix factorization in recognition of mammograms," in *Eighth International Conference on Graphic and Image Processing (ICGIP 2016), SPIE, 2017*, pp. 53–59.
- [8] H. Xu and J. Ma, "EMFusion: An unsupervised enhanced medical image fusion network," *Inf. Fusion*, vol. 76, pp. 177–186, 2021.
- [9] B. Lin, X. Tao, and J. Lu, "Hyperspectral Image Denoising via Matrix Factorization and Deep Prior Regularization," *IEEE Trans. Image Process.*, vol. 29, pp. 565–578, 2020, doi: 10.1109/TIP.2019.2928627.
- [10] S. Mirzaei, S. Khosravani, and others, "Hyperspectral image classification using non-negative tensor factorization and 3D convolutional neural networks," *Signal Process. Image Commun.*, vol. 76, pp. 178–185, 2019.

- [11] D. Li, Z. Gao, X.-P. Zhang, G. Zhai, and X. Yang, "Generative adversarial networks for non-negative matrix factorization in temporal psycho-visual modulation," *Digit. Signal Process.*, vol. 100, p. 102681, 2020.
- [12] P. Peng et al., "Group sparse joint non-negative matrix factorization on orthogonal subspace for multi-modal imaging genetics data analysis," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 19, no. 1, pp. 479–490, 2020.
- [13] S. Liu, M. Li, Z. Zhang, B. Xiao, and T. S. Durrani, "Multi-Evidence and Multi-Modal Fusion Network for Ground-Based Cloud Recognition," *Remote Sens.*, vol. 12, no. 3, p. 464, Feb. 2020, doi: 10.3390/rs12030464.
- [14] H. Liu, T. Fang, T. Zhou, and L. Wang, "Towards Robust Human-Robot Collaborative Manufacturing: Multimodal Fusion," *IEEE Access*, vol. 6, pp. 74762–74771, 2018, doi: 10.1109/ACCESS.2018.2884793.
- [15] J. Gao, Y. Lu, J. Qi, and L. Shen, "A Radar Signal Recognition System Based on Non-Negative Matrix Factorization Network and Improved Artificial Bee Colony Algorithm," *IEEE Access*, vol. 7, pp. 117612–117626, 2019, doi: 10.1109/ACCESS.2019.2936669.
- [16] S. V. Yakkundi and D. P. Subha, "Convolutional LSTM: A Deep learning approach for Dynamic MRI Reconstruction," in *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)*(48184), Tirunelveli, India: IEEE, Jun. 2020, pp. 1011–1015. doi: 10.1109/ICOEI48184.2020.9142982.
- [17] W. Ma et al., "Infrared and Visible Image Fusion Technology and Application: A Review," *Sensors*, vol. 23, no. 2, p. 599, 2023.
- [18] H. R. Almadhoun and S. S. A. Naser, "Detection of Brain Tumor Using Deep Learning," vol. 6, no. 3, p. 19, 2022.
- [19] Z. Huang, "Integrative Analysis of Multimodal Biomedical Data with Machine Learning," PhD Thesis, Purdue University Graduate School, 2021.
- [20] F. Gao, X. Deng, M. Xu, J. Xu, and P. L. Dragotti, "Multi-modal convolutional dictionary learning," *IEEE Trans. Image Process.*, vol. 31, pp. 1325–1339, 2022.
- [21] R. Soroush and Y. Baleghi, "NIR/RGB image fusion for scene classification using deep neural networks," *Vis. Comput.*, vol. 39, no. 7, pp. 2725–2739, Jul. 2023, doi: 10.1007/s00371-022-02488-0.
- [22] A. Khader, J. Yang, and L. Xiao, "NMF-DuNet: Nonnegative Matrix Factorization Inspired Deep Unrolling Networks for Hyperspectral and Multispectral Image Fusion," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 15, pp. 5704–5720, 2022, doi: 10.1109/JSTARS.2022.3189551.
- [23] M. Salvi, U. R. Acharya, F. Molinari, and K. M. Meiburger, "The impact of pre- and post-image processing techniques on deep learning frameworks: A comprehensive review for digital pathology image analysis," *Comput. Biol. Med.*, vol. 128, p. 104129, Jan. 2021, doi: 10.1016/j.combiomed.2020.104129.
- [24] H. Xu, J. Ma, J. Jiang, X. Guo, and H. Ling, "U2Fusion: A unified unsupervised image fusion network," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 1, pp. 502–518, 2020.
- [25] R. Dian, S. Li, B. Sun, and A. Guo, "Recent advances and new guidelines on hyperspectral and multispectral image fusion," *Inf. Fusion*, vol. 69, pp. 40–51, 2021.

# Hybrid Image Encryption using Non-Adjacent Bits Dynamic Encoding DNA with RSA and Chaotic Systems

Marwa A. Elmenyawi<sup>1</sup>, Nada M. Abdel Aziem<sup>2</sup>

Benha Faculty of Engineering, Benha University Benha, Egypt<sup>1</sup>  
Arab Academy for Science, Technology and Maritime Transport - Arab League<sup>1,2</sup>

**Abstract**—Image encryption is a crucial aspect that helps to maintain the images' confidentiality and security in diverse applications. Ongoing research is focused on improving the efficiency and effectiveness of encryption. Image encryption has many practical applications in today's digital world, such as securing confidential images transmitted over networks, protecting sensitive personal information stored in images, and ensuring the privacy of medical images. The suggested work represents a breakthrough in image encryption by proposing a model that leverages the power of DNA, RSA, and chaos. This model has three phases: key generation, confusion, and diffusion. The key generation phase employs a hash function and hyperchaotic technique to generate a strong key. During the confusion phase, the positions of pixels are rearranged, either at the image level or within blocks, using the Duffing chaotic map. Once the scrambling level is determined, each pixel undergoes two successive scrambling steps, with Henon and Arnold's chaotic map to change its location. During the diffusion phase, the encryption model employs a two- approach to ensure maximum security. Firstly, it utilizes dynamic DNA cryptography for non-adjacent bits, followed by robust RSA cryptography. The experimental results indicate that the model possesses a strong security level randomness and can withstand different attacks.

**Keywords**—Cryptography; image encryption; hash function; chaotic map; DNA encoding; DNA operations; RSA algorithm

## I. INTRODUCTION

In today's digital era, digital images are widely used for personal, professional, or commercial purposes. Therefore, these images require protection from unauthorized access. Image encryption and information hiding are the two basic approaches for securing digital images. Image encryption prevents hackers from recognizing the images by employing complex mathematical operations to transform image data into an unreadable form. Therefore, the hacker's attempts are wasted. There is a need to design and implement an algorithm characterized by its security and efficiency to succeed in withdrawing the different attacks. There are two phases in image encryption: diffusion and scrambling. The pixel positions are altered during the scrambling phase, whereas the pixel values are changed during the diffusion phase.

Various methods are employed during the confusion phase. Some of these works used the Arnold transform [1-3], Zigzag transformation [4,5], Fisher-Yates [6], and Josephus traversal

[7]. Other works implemented scrambling over two steps, such as L-shape and Arnold transforms as in [8], new filling curve design and Josephus traversal [9].

In terms of diffusion, S. Wang. et al. [10] suggested using the DNA sequence in the diffusion phase. Four sequences were derived from a 4D chaotic system and utilized to select the rules for encoding, computing, and decoding. They utilized higher dimensional chaos to offer a large key space. J. Yu et al. [11] began with the diffusion phase. They encoded the three matrices of RGB image using DNA sequence where a chaotic system chooses the rule. A new operation, known as DNA triploid mutation, was introduced to achieve cryptographic translation of DNA bases. Finally, they permuted the image using row-column permutation. C. Zou et al. [12] utilized two types of DNA strands: long and short. The image was permuted using two short DNA strands, while the long DNA strand was used in the diffusion stage. If the DNA sequence follows the property of the Watson-Crick base pairing, the XOR operation of DNA is performed; otherwise, the DNA addition operation is used.

B. Jasra and A. Moon [13] split the color image into three planes and encoded each plane using a DNA sequence based on the chosen row-level. A substitution algorithm relies on elliptic curves to accomplish effective encryption and authentication. J. Wang et al. [1] suggested a new type of chaos; Logistic-Sine self-embedding. They proved this type's chaotic features and adopted a 0-1 test to find the chaos's presence in the time series. They encrypted the plain image using a Logistic-Sine self-embedding chaotic system. Similarly, X. Li [14] introduced another chaotic sequence, which was 5D, and they showed that the 5D chaotic did not have a prominent Lyapunov exponent yet possessed several good characteristics. The 5D chaotic sequence was utilized to choose the DNA encoding, computing, and decoding.

The encryption model presented in [15] depended on a fused magic cube produced by fusing two magic cubes. The cipher image's pixels value was obtained from the plain image's pixels value by employing the fuse magic cube. J. Zheng and Q. Zeng [16] constructed an S-box using the obtained key from the Logistic map, generating a 16 x 16 matrix ranging from 0 to 255 with no repeated values. The image was diffused by traversing the scrambled image in order according to the generated keys.

A technique for encrypting color images was introduced in [17], which employs 3D chaos, RSA, DNA, and LSB. The image is initially encrypted using a DNA method and Lorenz chaotic map. The secret key is then encrypted employing RSA and hidden within a cipher image using LSB. In another approach proposed by M. Liu and G. Ye [18], image encryption is achieved by utilizing dynamic DNA alongside a hyperchaotic system. The dynamic DNA coding selects DNA rules in a randomized manner, guided by the employed chaotic map. This study also incorporates RSA to protect the secret key's confidentiality during transmission and management. U. Mir et al. [19] introduced an encryption method for color images using RSA and chaos in the domain of Hartley. The image is first ciphered utilizing RSA and then transformed from the time to frequency domain using a Hartley domain.

K. Jiao et al. [20] introduced an encryption approach combining RSA and a generalized Arnold chaotic map. The RSA algorithm obtains the map parameters, generating the keystream for a diffusion operation on the plaintext image. The confusion operation is then employed to conceal the image data and produce the cipher image. Babu M et al. [21] used chaotic Maps, RSA, and DNA sequences to encrypt images. This paper divided the plain image into several blocks. Secondly, different encryption schemes were utilized on each block, such as Secure Force, DNA Sequence, Arnold Map, and RSA encryption. Thirdly, a discrete cosine transformation algorithm was applied to the merged blocks, after which an XOR operation was conducted with a randomly generated key to produce the encrypted image.

This paper aims to enhance image encryption security by minimizing pixel correlation, maximizing randomness and unpredictability, and withstanding various types of attacks. The proposed encryption approach has three phases: key generation, confusion, and diffusion. The integration of different chaotic maps leads to a substantial expansion of the key space and makes encryption impervious to brute-force attacks. A robust key generation process achieves cryptosystem robustness against various attacks. Chaos and hash functions, SHA and MD5, produce the encryption key. The advantages of the SHA function are its irreversibility and a one-time pad key, while the chaos is characterized by randomness and unpredictability. The final encryption key is obtained by applying these hash functions to the user-specified key and the plaintext image. The encryption security approach is strengthened by utilizing the plaintext image and the user key.

Scrambling a pixel's locations can be done over the whole image or the divided blocks of the whole image, depending on the Duffing chaotic map. Moreover, two levels of confusion are implemented using Arnold and Henon's chaotic maps. The scrambling phase achieves high randomness between the pixels and decreases the correlation between the pixels to the minimum compared to the previous research, as illustrated by the results of the suggested method. The final phase is diffusion, implemented over two steps: DNA and RSA. In the DNA algorithm, the pair of bits to be replaced with the DNA sequence is not successive as customary in state-of-the-art research. The proposed algorithm incorporates dynamic DNA to select various rules for each pixel and DNA computations to improve its efficiency. The encoding, computation, and

decoding rules are determined using a 4D hyperchaotic sequence.

The second diffusion level is to implement the RSA algorithm, which is different from the state-of-the-art research where most of the paper used RSA along with DNA utilized RSA outputs as the initial values of chaotic sequences. Despite using multiple steps and various types of chaotic maps, we tried to keep the algorithm's runtime comparable to previous research. We conducted experimental testing and analysis to demonstrate the proposed approach's superiority and feasibility, showcasing its resilience against multiple attacks such as differential, plaintext, brute-force, occlusion, and noise attacks. Moreover, the correlation is lower than in the most recent research.

The paper's structure is as follows: the second section provides an overview of chaos and DNA cryptography. The third section outlines the suggested approach for image encryption/decryption, followed by the fourth section thoroughly explores the results, conducts in-depth analysis, and compares them with similar research. Lastly, the fifth section introduces the paper's conclusion.

## II. BACKGROUND

### A. Chaotic

Chaotic systems are characterized by unique features appropriate in encryption, like sensitivity to initial conditions, irregular behavior, and unpredictability. Therefore, using the chaotic sequence in the encryption system can give a high-security degree and robustness against attacks. Our suggested algorithm employs various types of chaotic systems at different stages to increase the key space and enhance security.

The 1D logistic map [22] generates chaotic dynamics in a discrete-time system. The Logistic chaotic map provides high speed, low arithmetic operations, and low computational overhead. It is a nonlinear recursive function defined by Eq. (1).

$$x_{n+1} = rx_n(1 - x_n) \quad (1)$$

Where  $r$  is a parameter that determines the map behavior.  $x_0$  should be  $\in [0,1]$  and  $r$  has to be within interval  $0 < r \leq 4$  to produce the chaotic behavior.

The Henon Chaotic Map [19] refers to a discrete-time dynamical map in 2D. Eq. (2) and (3) define the equations of the Henon map:

$$x_{n+1} = 1 - a x_n^2 + y_n \quad (2)$$

$$y_{n+1} = b x_n \quad (3)$$

To attain chaotic behavior in the Henon Chaotic Map, the control parameters  $a$  and  $b$  need to be assigned the values (1.4, 0.3).

The Arnold Chaotic Map [23] is often employed to scramble and alter pixel locations. It is described in Eq. (4) and (5).

$$x_{n+1} = (2x_n + y_n) \bmod m \quad (4)$$

$$y_{n+1} = (x_n + y_n) \bmod m \quad (5)$$

The "mod m" operation ensures that the coordinates do not exceed the image size.

Quantum Chaotic Map [24] is a classical dynamical system that is described to develop the function for solving computing of the quantum. It is used to generate many random numbers. The quantum map mathematical expression is given in Eq. (6)-(8).

$$x_{n+1} = r(x_n - |x_n|^2) - ry_n \quad (6)$$

$$y_{n+1} = -y_n e^{-2\beta} + e^{-\beta} r[(2 - x_n - x_n^*)y_n - x_n z_n^* - x_n^* z_n] \quad (7)$$

$$z_{n+1} = -z_n e^{-2\beta} + e^{-\beta} r[2(1 - x_n^*)z_n - 2x_n y_n - x_n] \quad (8)$$

Where  $x \in [0,1]$ ,  $y \in [0, 0.1]$ ,  $z \in [0, 0.2]$ ,  $r \in [0,4]$ ,  $\beta \in [6, \infty)$ ,  $x^*$  and  $z^*$  are complex conjugates of  $x$  and  $z$ .

Duffing Chaotic [23] produces chaotic dynamics in a discrete-time system. The Eq. (9) and (10) represent this map. It can be utilized to design damped oscillators.

$$x_{n+1} = y_n \quad (9)$$

$$y_{n+1} = -b x_n + a y_n - y_n^3 \quad (10)$$

Its control parameters,  $a$  and  $b$ , should be 2.75 and 0.2 to maintain chaotic behaviour in the Duffing Chaotic system.

### B. DNA

Images are encrypted using nucleic acid bases via a DNA encryption system, which subsequently carries out several additional DNA operations. The four nucleic acid bases identified by the Watson-Crick basic pairing principles are C (Cytosine), A (Adenine), T (Thymine) and G (Guanine). Every two bases complement each other; A is the complement of T and likely C and G. These four bases can be represented in binary using the numbers 00, 01, 10, and 11. There are 24 possible combinations for the DNA bases represented in binary. However, only eight satisfy the Watson-Crick complementary rule. Table I lists these eight coding principles. An illustration of the encoding process is as follows: Suppose a pixel value of 90 is represented in binary as 01011010. This number is then encoded as the AATT sequence if rule 4 is applied. There are a different number of operations according to the binary system. The operations used in this paper are addition, subtraction, multiplication, XNOR, XOR, right rotate and left rotate [10].

TABLE I. THE RULES OF DNA

Rule	1	2	3	4	5	6	7	8
A	00	00	01	01	10	10	11	11
T	11	11	10	10	01	01	00	00
C	01	10	00	11	00	11	01	10
G	10	01	11	00	11	00	10	01

## III. PROPOSED CRYPTOSYSTEM APPROACH

This section will showcase the construction of our suggested algorithm, which comprises three phases: key generation, confusion, and diffusion. The complete layout of our suggested algorithm is introduced in Fig. 1, and each phase will be elaborated in further detail in the preceding subsections. Our suggested approach is applicable to both grayscale and color images. For a color image with a size of  $M \times N \times 3$ , the three colors are separated into three matrices with an  $M \times N$  size, and each matrix is processed as a plain image.

### A. Key Generation

The immunity of the cryptosystem to various attacks is based on producing a solid key. The suggested algorithm employs the hash functions and chaotic sequence to generate the encryption key. The hash function has the advantage of its irreversibility and is a one-time pad key, while the chaos is characterized by randomness and unpredictability.

We chose to use MD5 and SHA-256 as the hash functions. The MD5 is faster than the SHA-256, but the SHA-256 is more complex than MD5. A hash value  $H_k$  is generated based on combining the original image and a random user-specified key. Generating the key from the original image enables the system to withstand chosen/known-plaintext attacks. The final key value,  $K_i$ , consists of 32 bits decimal.

The final key is used to generate the second input of DNA operation using the hyperchaotic map, Eq. (11)-(14). To enhance the approach's resistance to brute force attacks, we opted for the hyperchaotic map, which offers a large key space. The hyperchaotic initial values are calculated according to Eq. (15)-(18) and utilizing the final key,  $K_i$ . The matrix  $M \times N$  constitutes the representation of the second input, possessing identical dimensions to those of the original image.

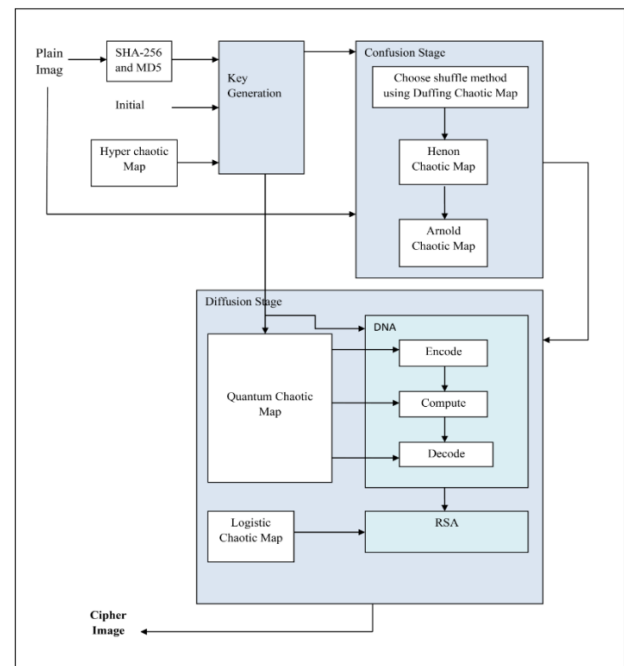


Fig. 1. The proposed image encryption approach's block diagram.

$$x_{n+1} = a(x_n - y_n) + r w_n \quad (11)$$

$$y_{n+1} = b x_n + x_n z_n - q y_n \quad (12)$$

$$z_{n+1} = -x_n y_n - c z_n \quad (13)$$

$$w_{n+1} = d x_n z_n - k w_n \quad (14)$$

$$x_0 = \frac{((((k_1 \oplus k_2) \wedge k_3) \oplus k_4) \wedge k_5) \oplus k_6}{256} \quad (15)$$

$$y_0 = \frac{((((k_7 \oplus k_8) \wedge k_9) \oplus k_{10}) \wedge k_{11}) \oplus k_{12}}{256} \quad (16)$$

$$z_0 = \frac{((((k_{13} \oplus k_{14}) \wedge k_{15}) \oplus k_{16}) \wedge k_{17}) \oplus k_{18}}{256} \quad (17)$$

$$w_0 = \frac{((((k_{25} \oplus k_{26}) \wedge k_{27}) \oplus k_{28}) \wedge k_{29}) \oplus k_{30}}{256} \quad (18)$$

Where  $k_i$  is XOR between the generated hash value and the initial secret key.

### B. Confusion Stage

This stage aims to change the pixel locations. Two alternatives were presented to achieve this goal: either process each pixel across the entire image or partition the image into blocks and rearrange the position of pixels within each block. The option number is calculated by Eq. (19), where  $y$  is determined from Duffing chaos in Eq. (9) and (10). Eq. (20)-(23) are utilized to compute the Duffing map's initial values and control parameters.

$$sh_{n0} = \text{floor}(\text{mod}(\text{avg}(y), 2)) \quad (19)$$

$$x_0 = \frac{k_{19} \oplus k_{20} \oplus k_{21} \oplus k_{22} \oplus k_{23}}{16} \quad (20)$$

$$y_0 = \frac{k_{25} \oplus k_{26} \oplus k_{27} \oplus k_{28} \oplus k_{29} \oplus k_{30} \oplus k_{31}}{64} \quad (21)$$

$$a = \frac{(k_{15} \oplus k_{16}) \oplus (k_{17} \wedge k_{18})}{16} \quad (22)$$

$$b = \frac{k_1 \oplus k_2 \oplus k_3 \oplus k_4 \oplus k_5 \oplus k_6}{512} \quad (23)$$

If the option number is one, we will change the pixel locations on the image's level depending on two chaotic systems: Arnold and Henon maps. The first step is to convert the original location,  $i$  and  $j$ , of each pixel to  $i_h$  and  $j_h$  locations using Henon, as shown in Eq. (24) and (25). The final locations,  $i_f$  and  $j_f$ , are generated using Arnold chaotic using the generated locations from the previous step, as illustrated in Eq. (26) and (27).

$$i_h = \text{mod}(\text{round}(|1 - a i^2 + j|), M) + 1 \quad (24)$$

$$j_h = \text{mod}(\text{round}(i + c), N) + 1 \quad (25)$$

$$i_f = \text{mod}(i_h - 1 + j_h - 1, M) + 1 \quad (26)$$

$$j_f = \text{mod}(i_h - 1 + 2(j_h - 1), N) + 1 \quad (27)$$

The other alternative is to modify the pixel location at the block level. The initial step involves partitioning the image into four blocks and shuffling their positions. Then, the pixel location within the block is rearranged by applying the same procedures as in the first method.

### C. Diffusion Stage

The confusion stage alone did not meet the encryption security requirements, necessitating the addition of a diffusion stage. The diffusion stage effectively conceals the original image information and increases attack resistance. In our proposed system, this stage consists of two steps, DNA and RSA, to offer more security to the cryptosystem.

1) DNA: In the DNA step, the encryption algorithm converts each pixel into its corresponding binary representation. Then, every pair of bits is substituted with a DNA sequence of four bases based on one of the eight DNA encoding rules shown in Table I. The DNA rule choice is made dynamically and calculated from Eq. (28), depending on the quantum map. The generated key is employed to obtain the quantum sequence's initial values, as shown in Eq. (29)-(31). The number of the generated quantum sequences equals the number of pixels to confuse the attacker, which rule is chosen and makes the cryptographic system unpredictable. Unlike the previous research, we did not replace the adjacent bits in each pixel; instead, we constitute a pair from the bit  $i$  and bit  $i+2$ , not  $i+1$ .

$$R_i = \text{floor}(8|x_i|) + 1 \quad (28)$$

$$x_0 = \frac{(((((((k_1 \oplus k_2) \oplus k_3) \oplus k_4) \oplus k_5) \oplus k_6) \oplus k_7) \oplus k_8) \oplus k_9)}{512} \quad (29)$$

$$y_0 = \frac{(((((((k_{10} \oplus k_{11}) \wedge k_{12}) \oplus k_{13}) \wedge k_{14}) \oplus k_{15}) \wedge k_{16}) \oplus k_{17}) \wedge k_{18})}{512} \quad (30)$$

$$z_0 = \frac{(((((((k_{19} \oplus k_{20}) \wedge k_{21}) \oplus k_{22}) \wedge k_{23}) \oplus k_{24}) \wedge k_{25}) \oplus k_{26}) \wedge k_{27})}{512} \quad (31)$$

$$op_i = 7 * \text{floor}(10^8|z_i|) \quad (32)$$

An illustration of the encoding process is as follows: Suppose a pixel value of 90 is represented in binary as 01011010. This number is then encoded using rule 4 as the GCCG sequence, not AATT. Here, the first A is determined using the first and third bit, which gives pair 11, equivalent to A according to rule 4.

After the encoding step, we apply a DNA operation on the encoded DNA sequence and the generated input from the previous stage to obtain another DNA sequence. The DNA operation,  $op_i$ , is selected based on the quantum sequence, as shown in Eq. (32). Finally, we perform the DNA decoding, which is the encoding reverse. The DNA sequence is converted to its equivalent binary using the rules in Table I. The rule is selected according to the quantum sequence shown in Eq. (28). The corresponding binary bits are reordered in odd and even positions, not as successive bits. For example, if the sequence is ATCG and the used rule is 8, the binary equivalent is 11001001, then it is reordered to 10101001. Finally, the decimal equivalent of the binary number is produced to use as the input of the RSA step.

2) RSA: The RSA step involves generating random prime numbers  $p$  and  $q$  using a Logistic chaotic map. The public and private keys are then calculated based on the result generated from the Logistic chaotic map, as shown in Algorithm (1). Afterwards, image encryption is achieved using the public



key, and decryption using the private key is illustrated in Algorithm (2).

---

**Algorithm (1) Public and private key generation**

---

Input: The prime numbers (p, q)  
 Output: Public key (PU), Private key (PR)  
 $x=q*p$   
 $\Phi(x)=(q-1)*(p-1)$   
 Choose e in condition that  $1 < e < \Phi(x)$  and  $\gcd(\Phi(x), e) = 1$   
 $d \equiv e^{-1} \pmod{\Phi(x)}$   
 PU = {x, e}  
 PR = {x, d}

---

**Algorithm (2) Encryption/Decryption of the image**

---

Input: Public key (PU), Private key (PR)  
 Output: Cipher image (C), Plain image (P)  
 $C = P^e \pmod{x}$   
 $P = C^d \pmod{x}$

---

**D. Image Decryption**

In the decryption phase, the encrypted image is converted back to its initial form, and the decrypted image reproduces the original image. This process follows a pattern inverse to that of encryption. The encrypted image first undergoes the diffusion phase, utilizing RSA followed by DNA techniques. Next, it enters the confusion phase, where Arnold and Henon chaotic maps are used, and the key generated from chaotic and hash functions, MD5 and SHA-256, is applied. The final output of this process is the original image.

**IV. EXPERIMENTAL RESULTS AND ANALYSIS**

The effectiveness and robustness of the suggested algorithm through various security tests are demonstrated in this section. Using MATLAB R2021a, the algorithm was simulated on a computer with an Intel(R) Core (TM) i5-6200U CPU @ 2.3GHz 2.4GHz and 8 GB of memory. There are different colors and grey images with different sizes utilized for testing. The grey images used in testing are Male 1024x1024, Lena 512x512, Barbara 512x512, Lake 512x512, Cameraman 256x256, and Kitten 256x256. The color images are Shreveport 1024x1024, Lena 512x512, Baboon 512x512, Lena 256x256, and Couple 256x256. Most of the images used in the study were sourced from the USC-SIPI database. Illustrations depicting the plaintext, cipher, and decrypted images are presented in Fig. 2 and Fig. 3 (a, c, and e). The right keys can be utilized to restore all images during the decryption tests. Any change in the secret keys, even if slight, leads to incorrect image decrypting, as will be proved in the following subsections. According to [25], there are four categories to estimate the proposed algorithm's performance.

- The visual perception evaluation

This category aims to generate an uncorrelated cipher image from the plain image. The performance metrics included in this category are PSNR, MSE, Correlation, Entropy, and Histogram analysis. Another test carried out within this category is the similarity test using the SSIM performance, which is adopted to evaluate the matching degree between the plaintext and cipher images.

- The high-performance evaluation

The goal of this category is to evaluate the diffusion characteristics. The tests to accomplish this goal are NPCR and UACI.

- The processing time.
- The strength of the cryptosystem evaluation

The cryptosystem strength is measured through key space, key sensitivity, and attack resistance.

**A. The Visual Perception Evaluation**

1) *Histogram analysis*: The distribution of tones inside an image, made up of pixels with various grey values, is essential. The histogram, which can adequately depict the amount of each pixel value, represents the tonal distribution of an image. An image's histogram gives statistical information that can be used to assess how strong an encryption system stands up to statistical attacks. Any plain image's histogram is covered by an angled or curved bar. There is much information in this bar. Hackers can use this bar to further their harmful goals. The cipher's task is to modify the pixel intensity values to produce a histogram with a uniform bar above it. Any information leaking in the image is greatly discouraged by the histogram's uniformity of the bar.

Additionally, it intimidates hackers and prevents them from succeeding with histogram attacks. Fig. 2 and 3(b,d) introduces the plain and encrypted images histogram. Apparently, the pixel distribution in cipher images is relatively uniform across all channels, but plain images have several peaks.

2) *PSNR and MSE*: Several measures are used to evaluate image quality, including PSNR and MSE, which quantify the level of difference between the original and cipher images. Equations (33-34) provide the mathematical expressions for PSNR and MSE.

$$PSNR = 10 \times \log_{10} \frac{p^2}{MSE} \quad (33)$$

$$MSE = \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N |O(i,j) - C(i,j)|^2 \quad (34)$$

The maximum pixel value, represented by 8 bits, is denoted by  $p = 255$ . The plaintext and cipher images are denoted by  $O(i,j)$  and  $C(i,j)$ . The higher MSE and lower PSNR values determine the method's efficiency and security. Table II presents the PSNR and MSE values for grayscale and color image channels. As elaborated in Table II, the suggested algorithm achieves low PSNR and high MSE values, indicating its strong security and efficiency. Moreover, Table III compares the suggested algorithm's performance with previous research, pointing to its better performance relative to other techniques concerning both PSNR and MSE.

3) *SSIM*: SSIM is a test used to indicate how much the cipher image is similar to the plaintext image depending on the amount of structural information modification of the plaintext image. The SSIM is obtained as indicated in Eq. (35). A decrease in the similarity index indicates a decrease in

the match between the original and encrypted images and an increase in the degree of changed structural information.

The calculated values of the SSIM index are introduced in Table II. Table III shows that lower SSIM values close to zero indicate the lower similarity of the suggested algorithm. Moreover, the suggested algorithm gives better results than the other research, meaning the proposed algorithm offers less similarity to the other research, as indicated in Table III.

$$\begin{cases} SSIM(O, E) = \frac{(2 \mu_O \mu_E + C_1)(2 \sigma_{OE} + C_2)}{(\mu_O^2 \mu_E^2 + C_1)(\sigma_O^2 \sigma_E^2 + C_2)} \\ C_1 = (K_1 L)^2 \\ C_2 = (K_2 L)^2 \end{cases} \quad (35)$$

Where the  $\mu_O$  and  $\mu_E$  are the original and cipher images mean.  $\sigma_O^2$  and  $\sigma_E^2$  are the plaintext and cipher image variance. The original and cipher images covariance is denoted by  $\sigma_{OE}$ . L is set to 255, while  $K_1$  and  $K_2$  are set to 0.01 and 0.03, respectively.

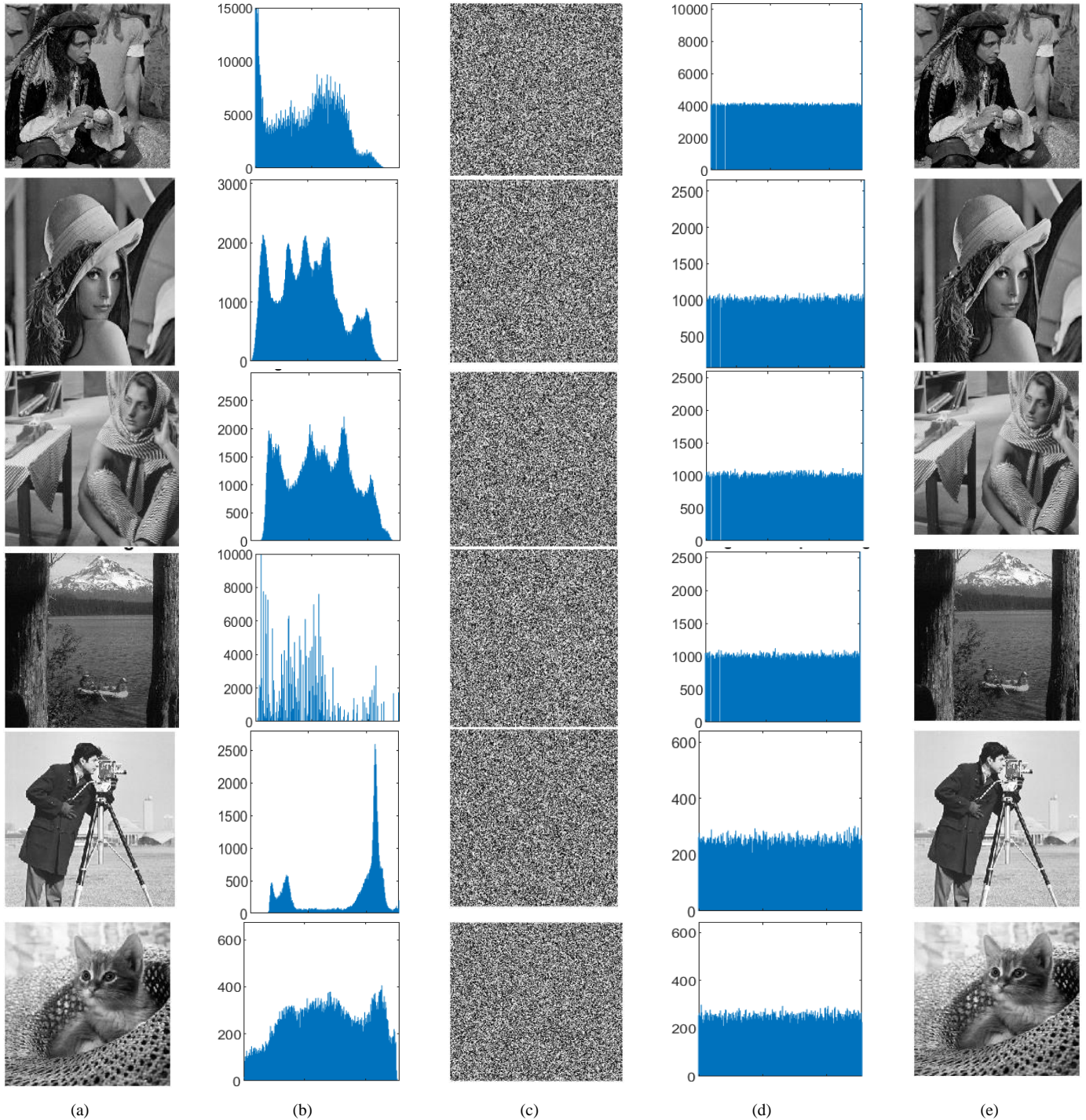
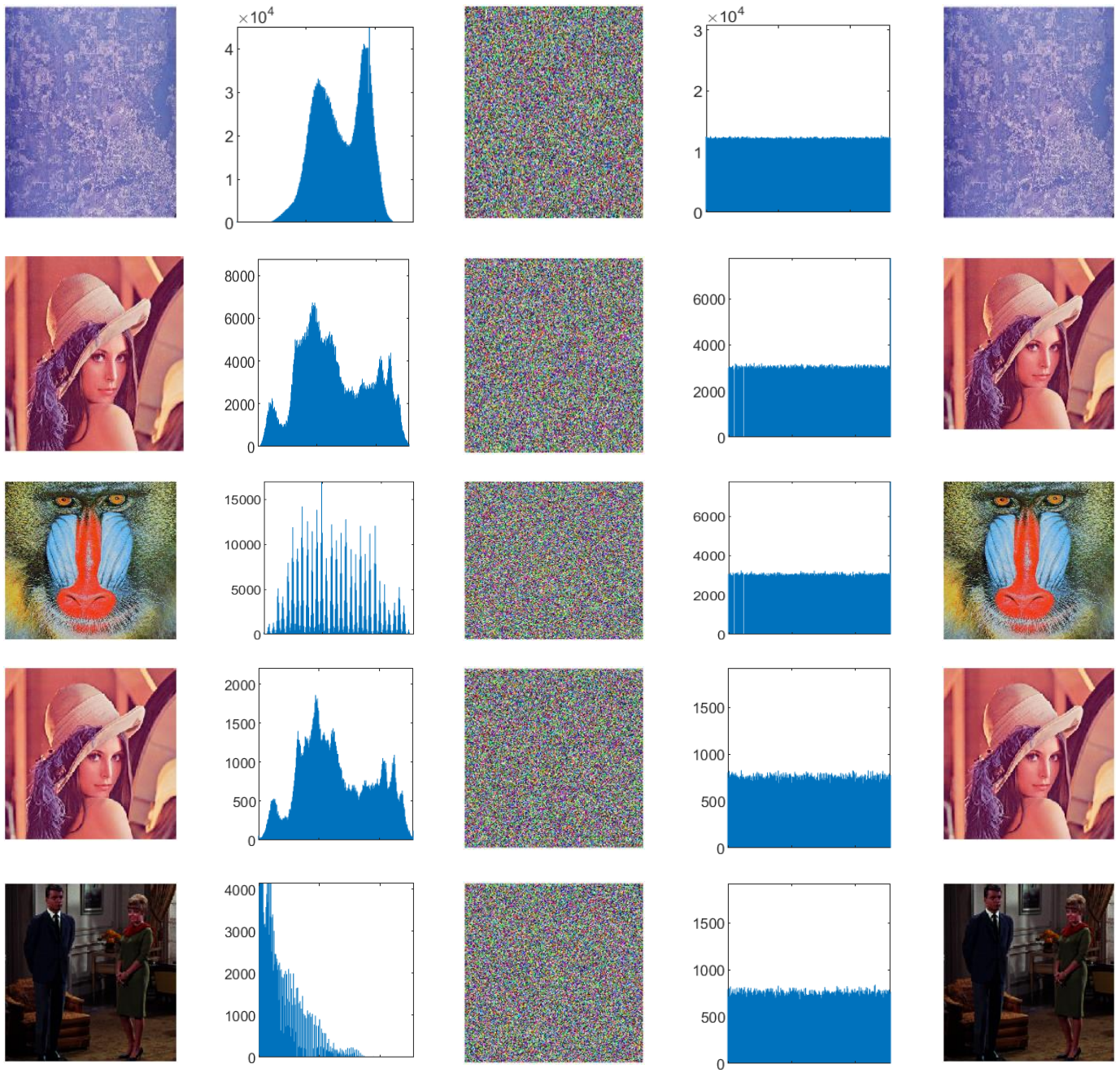


Fig. 2. The results of grayscale images (a) plain image (b) plain image histogram (c) cipher image (d) cipher image histogram (e) decrypted image.



(a) (b) (c) (d) (e)  
Fig. 3. The results of color images (a) plain image (b) plain image histogram (c) cipher image (d) cipher image histogram (e) decrypted image.

TABLE II. THE MSE, PSNR, MAE, AND SSIM PERFORMANCE

Image	Color	MSE	PSNR	SSIM
Male 1024 x1024	Grey	10475	7.9293	0.0001
Barbara512 x 512	Grey	8589.41	8.7912	0.0002
Lake 512 x 512	Grey	10893.21	7.7592	-0.0004
Kitten 256 x256	Grey	9685.53	8.2696	0.0005
Cameraman256 x 256	Grey	11675.60	7.4580	-1.3E-05
Shreveport 1024 x 1024	R	6463.01	10.0265	0.0005
	G	6151.77	10.2408	-0.0002
	B	9052.82	8.5630	-0.0001
Baboon 512 x 512	R	8627.04	8.7722	0.00097
	G	7902.03	9.1534	0.0006
	B	9950.34	8.1524	0.0003
Lena 256 x 256	R	10666.34	7.8506	-0.0006
	G	9048.21	8.5652	-0.0004
	B	7025.62	9.6640	0.0030
Couple 256 x 256	R	14092.71	6.6409	-0.0007
	G	15923.85	6.1103	-4.40354E-05
	B	16261.35	6.0192	0.0002

4) *Correlation analysis:* Correlation analysis is a measure of how closely two variables are related. In plaintext images, adjacent pixels tend to be highly correlated, which can be utilized to attack the image. If the neighboring pixels correlation is excessively high, it makes it easier for attackers

to predict the next pixel value. By breaking the correlation between pixels, statistical attacks can be prevented. As the correlation between pixels approaches zero, it becomes progressively more challenging for a potential attacker to deduce any insights into the original plaintext image. The correlation values mathematical formulas in three directions are computed by Eq. (36). In Table IV, the correlation coefficients between the original and cipher images are depicted for all three directions. As depicted in the table, the original image coefficients are near one, meaning a solid correlation between pixels.

$$\begin{cases} R_{xy} = \frac{cov(x,y)}{\sqrt{V(x)V(y)}} \\ cov(x,y) = \frac{1}{N} \sum_{i=1}^N (x_i - E(x))(y_i - E(y)) \\ V(x) = \frac{1}{N} \sum_{i=1}^N (x_i - E(x))^2 \\ E(x) = \frac{1}{N} \sum_{i=1}^N x_i \end{cases} \quad (36)$$

Where the pixel total number is denoted by N, the values of the neighboring pair are x and y. The variance, mean, and covariance are V(x), E(x), and Cov(x,y), respectively.

The correlation values within the encrypted images approach zero, indicating a minimal correlation among pixels. Moreover, Table V shows that the correlation values for the proposed algorithm outperform the previous methods, which satisfies our objective of minimizing the correlation. In Fig. 4, the correlation distribution among adjacent pixels is depicted for the three directions of the 512 x 512 color Lena image. The figure shows the equal cipher image distribution, meaning the correlation between pixels is low. Moreover, scrambling over pixels on the block's level gives better results than scrambling pixels on the image's level, as introduced in Table VI.

TABLE III. COMPARISON OF MSE, PSNR, AND SSIM

Image	Ref.	MSE			PSNR			SSIM		
Grey Lena 512	Ours	9236.99			8.4755			-0.000208231		
	[26]	7797.7			9.2111			0.0350		
		R	G	B	R	G	B	R	G	B
Color Lena 512 x 152	Ours	10492.31	9218.75	7207.19	7.9221	8.4841	9.5531	0.0008	0.0005	0.0003
	[26]	10,637			7.8625			0.0331		
	[17]	8828.6			8.6719			0.0200		

TABLE IV. CORRELATION COEFFICIENTS

Image		Plaintext Image			Encrypted Image		
		V	H	D	V	H	D
Male	Grey	0.9813	0.9774	0.9671	-0.0002	-0.0003	-0.0008
Barbara		0.9589	0.8954	0.8830	0.0021	-0.0002	-0.0004
Lake		0.9679	0.9545	0.9395	-0.0001	-0.00056	0.0012
Kitten		0.9228	0.9505	0.8840	-0.00095	-0.0031	-5.6E-05
Cameraman		0.9549	0.9196	0.8962	-0.0015	0.0033	-0.0078
Shreveport	R	0.8231	0.8295	0.7621	-0.00086	-1.34999E-05	0.00046

	<i>G</i>	0.8207	0.8245	0.7621	0.0006	-0.0002	0.0004
	<i>B</i>	0.8617	0.8646	0.8188	-0.00057	0.00086	0.0002
<b>Baboon</b>	<i>R</i>	0.8595	0.9105	0.8474	-0.0001	-0.0003	0.0001
	<i>G</i>	0.7755	0.8594	0.7434	0.0004	0.00075	0.0037
	<i>B</i>	0.8697	0.8953	0.8296	0.0007	-0.0006	-0.0013
<b>Couple</b>	<i>R</i>	0.9562	0.9493	0.9176	0.00099	0.0005	0.0084
	<i>G</i>	0.9534	0.9308	0.9002	-0.0058	-9.03962E-05	0.0023
	<i>B</i>	0.9442	0.9178	0.8880	-0.0004	-0.0053	-0.0001

TABLE V. COMPARISON OF CORRELATION COEFFICIENTS

		<b>V</b>			<b>H</b>			<b>D</b>		
<b>Grey Lena 512 x 512</b>	<i>Ours</i>	0.0001			-0.0004			0.0014		
	[26]	-0.0113			-0.0215			0.0089		
	[18]	0.0014			-0.0011			s0.0043		
		<i>R</i>	<i>G</i>	<i>B</i>	<i>R</i>	<i>G</i>	<i>B</i>	<i>R</i>	<i>G</i>	<i>B</i>
<b>Color Lena 512 x 512</b>	<i>Ours</i>	-0.0009	0.0006	0.00002	-0.00007	0.0007	0.0014	0.0002	0.0006	0.0029
	[26]	-0.0027			0.0007			-0.0104		
	[17]	0.0197			-0.0043			0.0032		
<b>Color Lena 256</b>	<i>Ours</i>	0.0004	0.0003	-0.0048	0.0004	0.0036	0.0033	0.0004	0.0012	-0.0006
	[11]	0.0098	0.0000	-0.0004	-0.0044	-0.0013	-0.0061	-0.0013	0.0042	-0.0093
	[28]	0.0019			0.0020			-0.0025		
	[27]	0.003	-0.004	-0.0008	0.0003	0.001	-0.0009	0.0008	0.002	0.002
	[29]	0.0063	-0.0023	0.0087	-0.0015	0.0035	0.0053	0.0043	-0.0081	0.0011
	[19]	-0.0002	-0.0051	0.0016	0.0019	0.0024	0.0007	0.0008	0.0018	-0.0019

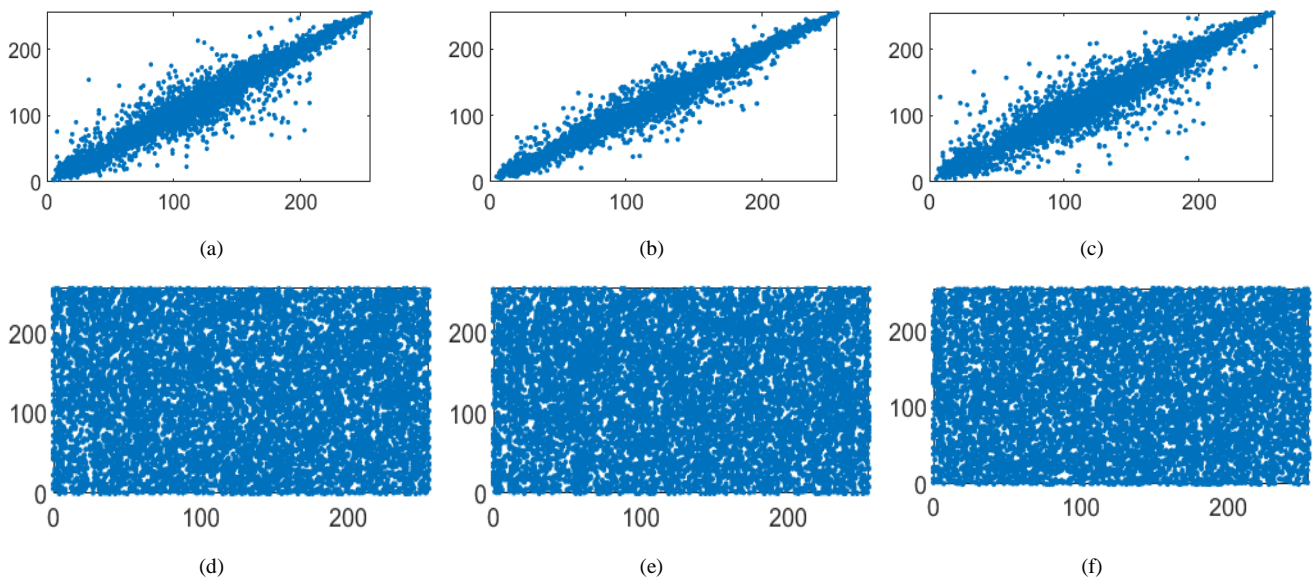


Fig. 4. Distribution of correlation values between original and cipher images in different directions for 512 x 512 color Lena image (a) original image's horizontal correlation (b) original image's vertical correlation (c) original image's diagonal correlation (d) cipher image's horizontal correlation (e) cipher image's vertical correlation (f) cipher image's diagonal correlation.

TABLE VI. COMPARISON BETWEEN THE SCRAMBLING ON THE IMAGE'S LEVEL AND THE BLOCK'S LEVEL

		V			H			D		
		R	G	B	R	G	B	R	G	B
Grey Lena 512 x 512	Block	0.0001			-0.0004			0.0014		
	Pixel	0.0014			0.0015			-0.0024		
Color Lena 512 x 512	Block	-0.0009	0.0006	0.00002	-0.00007	0.0007	0.0014	0.0002	0.00055	0.0029
	Pixel	0.0015	-0.0024	-0.0009	-0.0009	-0.0009	-0.0013	0.00198	0.0022	-0.0006
Color Lena 256	Block	0.0004	0.0003	-0.0048	0.0004	0.0036	0.0033	0.00036	0.0012	-0.0006
	Pixel	-0.008	0.0051	-0.0063	-0.0100	0.0021	-0.00085	-0.0002	-0.0038	-0.0024

5) Entropy: Entropy is a metric that indicates an image's level of randomness and unpredictability. The entropy can be calculated as given in Eq. (37).

$$Entropy = - \sum_{i=0}^{2^n-1} P(x_i) \log_2(P(x_i)) \quad (37)$$

Where the probability of x is P(x).

An entropy value of approximately eight is considered the ideal value for a cipher image [30]. Table VII presents the entropy value of the suggested algorithm, while Table VIII compares it with the recent work. The comparison reveals that the proposed method gives a better or similar result than the recent work, which satisfies our goal of maximizing the randomness and unpredictability of the image.

TABLE VII. THE ENTROPY, NPCR, AND UACI VALUES FOR THE SUGGESTED ALGORITHM

Image	Color	Entropy	NPCR	UACI
Male	Grey	7.99985	99.60	33.58
Barbara	Grey	7.99936	99.60	33.52
Lake	Grey	7.99936	99.60	33.56
Kitten	Grey	7.99752	99.62	33.48
Cameraman	Grey	7.99785	99.60	33.48
Shreveport	R	7.99983	99.61	33.48
	G	7.99984	99.61	33.52
	B	7.99983	99.60	33.48
Baboon	R	7.99926	99.62	33.60
	G	7.99922	99.61	33.62
	B	7.99933	99.60	33.63
Couple	R	7.99715	99.62	33.57
	G	7.99744	99.58	33.56
	B	7.99741	99.60	33.59

### B. The High-Performance Evaluation

In a differential attack, an attacker aims to uncover the differences between encrypted images produced from two slightly varied versions of the original image. The attacker looks for non-random areas in the encrypted images and then looks for changes in these areas that would conclude the key

used in image encryption. In slight alterations to the original image, the encryption procedure produces two cipher images: one for the original image and a second for the modified version. If a single-bit change in the original image causes the encrypted images to differ by at least 50%, then the differential attack cannot decrypt the cipher images. Two performance metrics are utilized, UACI and NPCR, to estimate the algorithm's resistance to differential attacks. The mathematical expressions for these two metrics are in Eq. (38) and (39).

$$NPCR = \frac{\sum_{x,y} D(x,y)}{N \times M} \times 100\% \quad (38)$$

$$UACI = \frac{\sum_{x,y} |E_1(x,y) - E_2(x,y)|}{255 \times M \times N} \quad (39)$$

Where  $E_1(x, y)$  and  $E_2(x, y)$  are two encrypted images of the same original image, but only one pixel value is changed.  $D(x,y)$  is calculated as follows:

$$D(x, y) = \begin{cases} 0 & E_1(x, y) = E_2(x, y) \\ 1 & E_1(x, y) \neq E_2(x, y) \end{cases}$$

Table VII showcases the NPCR and UACI values obtained through the proposed algorithm, whereas Table VIII provides a comparative evaluation of its performance against state-of-the-art methods. The NPCR and UACI values, approaching 99.6094% and 33.4635%, respectively, are considered close to the ideal benchmarks. The two metric values achieved by the suggested algorithm are in close proximity to the ideal values, demonstrating the suggested method's robustness against differential attacks.

### C. The Processing Time

Time evaluation is necessary in assessing the efficacy of cryptographic systems, with efficient systems expected to exhibit minimal encryption processes. To assess the time efficiency of our suggested algorithm, we conducted time measurements for each encryption step of the Kitten image, as shown in Fig. 5. The most consumable time is the key generation stage, as introduced in Fig. 5. Table IX displays a comparison of the encryption time consumed by the suggested approach with that of other approaches. The results show that our suggested method gives better time encryption for Lena with size 512. However, in the case of Lena's image with size 256, the results are higher than the two works as these works did not treat each color channel as a separate matrix as in our case.

TABLE VIII. THE ENTROPY, NPCR, AND UACI COMPARISON WITH PREVIOUS WORK

		Entropy			NPCR			UACI		
		R	G	B	R	G	B	R	G	B
Grey Lena 512 x 512	Ours	7.9992			99.61			33.67		
	[26]	7.9993			99.61			33.701		
	[18]	7.9994			99.61			33.47		
Color Lena 512 x 512	Ours	7.9993	7.9992	7.9994	99.60	99.59	99.60	33.63	33.52	33.53
	[13]	7.9924			99.61			33.78		
	[26]	7.9994			99.63			33.03		
	[17]	-----			99.62			30.45		
Color Lena 256	Ours	7.9971	7.9973	7.9976	99.64	99.63	99.66	33.69	33.49	33.49
	[11]	7.9968	7.9973	7.9974	99.60	99.61	99.61	33.53	33.38	33.67
	[10]	7.9974	7.9975	7.9973	99.63	99.62	99.62	33.51	33.32	33.46
	[27]	7.9892	7.9902	7.9896	99.61			32.95		
	[29]	7.9973	7.9973	7.9973	99.61	99.60	99.60	33.48	33.46	33.36
	[19]	7.9956	7.9954	7.9962	100	100	100	33.45	33.43	33.56

TABLE IX. COMPARISON OF ENCRYPTION TIME

	Our	[27]	[29]	[19]	[11]	[26]	[17]
Lena 256	2.318	4.455	0.375	0.282			
Lena 512	6.968	14.966	-----	-----	12.117	2.0179	80.05

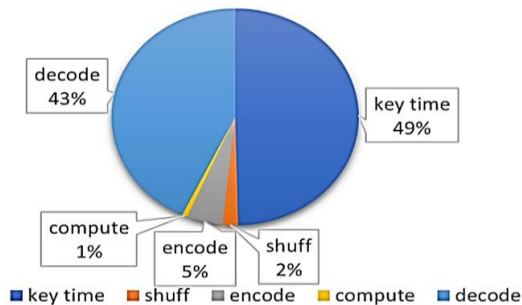


Fig. 5. The execution time of each stage in the encryption approach for the Kitten image.

#### D. The Strength of the Cryptosystem Evaluation

1) *Key space analysis*: There are different types of attacks to obtain the original image. One type of attack that hackers can use is Brute force, where they try to decrypt the cipher image by employing all possible keys until the correct one is found. A key space with a large size can be a good defense against brute-force attacks, which refers to the entire set of keys used for image encryption. The researchers have determined that a minimum key space of 2100 [31] is required to resist brute-force attacks. The image cryptosystem's secret key was generated using 24 parameters from different chaotic maps in the suggested method—additionally, the hash function used 256 bits.

Following the IEEE 754 floating-point standard (double), the substantial precision amounts to 53 bits, necessitating 15 decimal digits for representation. Consequently, the key space of the proposed system stands at  $10^{437}$ , demonstrating resilience against brute-force attacks by surpassing the threshold of  $2^{100}$ .

2) *Key sensitivity*: A cryptosystem is considered effective if it is extremely sensitive to keys. Key sensitivity means any minor change in key value results in a significant change in output. There are two ways to test key sensitivity. The first is by making a faint change in the key value; the result should be two different encrypted images. The other way is that the slight change in key-value results in not retrieving the original image correctly from the encrypted one. In this paper, we test the key sensitivity in two ways. In the first test, we changed the key slightly and then encrypted the plain image to get another cipher image using two different keys. We change the initial value of x and y in the hyperchaotic used to generate the final key by adding to each value  $10^{-14}$ . The results illustrated in Fig. 6 show the robustness of the suggested algorithm regarding the slight change in key as parts c and e show the difference between the cipher image generated from the valid key and the cipher image produced from the modified keys.

The alternative approach for conducting the key sensitivity test encompasses encrypting the original image with the accurate key and decrypting the resulting cipher image using an altered key. Fig. 7 illustrates the distinction between the decrypted image derived from the correct key and the decrypted image obtained from the two adjusted keys. The difference between the two images proves the suggested algorithm's high sensitivity against a minor key change. Based on the two test results, the suggested method can resist brute-force and statistical attacks.

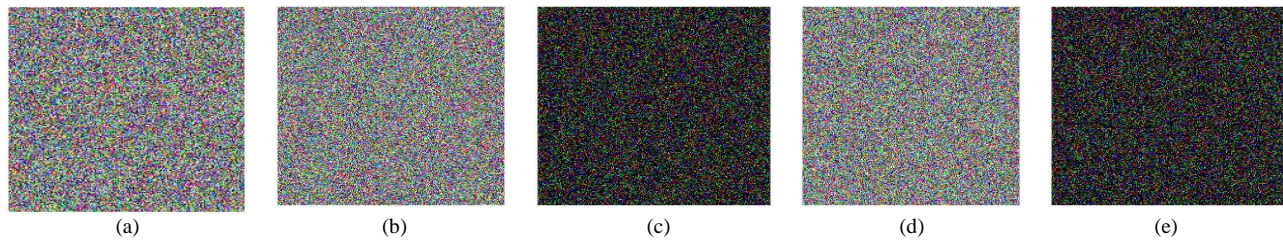


Fig. 6. Testing key sensitivity in encryption stage (a) image ciphered by the original key (b) image ciphered by the first modified key  $1(x0+10^{-14})$  (c) difference between a and b (d) image ciphered by the second modified key  $2(y0+10^{-14})$  (e) difference between c and d.

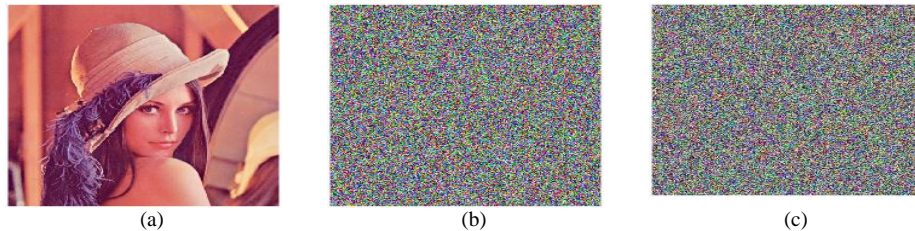


Fig. 7. Testing key sensitivity test in decryption (a) original image (b) decrypted image utilizing the first modified key  $(x0+10^{-14})$  (c) decrypted image utilizing the second modified key  $(y0+10^{-14})$ .

3) *Classical attacks*: Classical attacks include four attack types: chosen-plaintext, known-plaintext, chosen-ciphertext and ciphertext-only attack. The most harmful attacks are chosen and known plaintext attacks. In these two attacks, the attacker has access to plaintext and encrypted images and tries to deduce the keys. If the cryptographic system can withstand these attacks, it would be immune to the other two types.

The cryptographic system would be immune to the known and chosen plaintext attacks if it is sensitive to the key change, which is proved in subsection 2 of section D. Moreover, the suggested system is a one-time pad system that utilizes the SHA-256 function to produce the key. Another test in measuring the immunity to the attacks of known and chosen plaintext is to use all black and all-white images as the original, as shown in Fig. 8 (a) and (d). The corresponding cipher images for these two images and their histogram are displayed in Fig. 8 (b-c) and (e-f). The entropy and correlation values of the black and white images are depicted in Table X. From the results, the suggested algorithm shows strong immunity to the attacks of chosen and known plaintext attacks.

4) *Image processing attacks resistance*: During transmission, the cipher image may be exposed to several disruptions, and some attacks on cipher images will result in data loss. The decryption of a corrupted cipher image may thus lead to distorted or even unnoticeable results. The most famous image processing attacks are data and noise loss attacks. Minimizing the impact of data and noise loss attacks on the restored image is crucial for achieving an efficient image encryption algorithm.

The effectiveness of the proposed algorithm in withstanding noise attacks was assessed by introducing various levels of salt and pepper noise (0.5, 0.05, and 0.005) and then calculating the PSNR values between the original images and their decrypted counterparts. The results, presented in Table XI, affirm that the proposed algorithm effectively resists salt and pepper attacks.

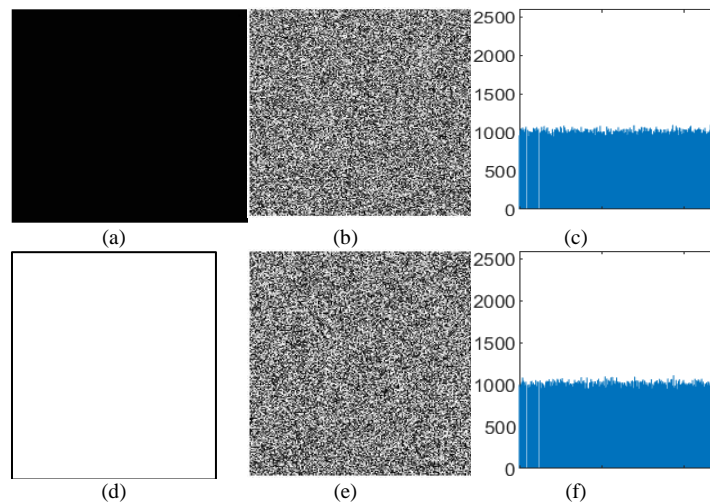


Fig. 8. Classical attack test(a-c) black image, its cipher, and its cipher histogram (d-f) white image, its cipher, and its cipher histogram.

TABLE X. THE ENTROPY AND CORRELATION OF THE WHITE AND BLACK IMAGES

image		Entropy	Correlation		
			V	H	D
White	plain	-0.000000	-	-	-
	cipher	7.9993	-0.00005	-0.0034	-0.00199
Black	plain	0.0012	-	-	-
	cipher	7.99938	0.00088	0.0011	-0.0033

Furthermore, the algorithm's ability to withstand occlusion attacks was assessed by applying varying masks to the cipher images (1/16, 1/8, and 1/4). The decrypted images resulting from this masking process are displayed in Table XI, showcasing the algorithm's robustness against cropping attacks. Despite occlusion attempts, the algorithm demonstrates its capability to recover a portion of the image's information.



TABLE XI. THE PSNR VALUE OF THE PLAINTEXT IMAGES AND THEIR DECRYPTED IMAGES AFTER APPLYING NOISE AND DATA LOSS ATTACKS

Attack	Parameters	PSNR		
		R	G	B
Salt and pepper	0.005	27.3081	27.8380	27.8922
	0.05	20.3436	20.8994	21.9327
	0.5	10.8365	11.5345	12.5723
Cropping	1/16	19.1520	20.9261	21.9408
	1/8	16.3463	17.9376	18.8688
	1/4	13.4978	14.8659	15.8419

### V. CONCLUSION

The suggested technique in this paper employs DNA, RSA, and chaotic maps to produce an extremely robust and secure image encryption method. The approach encompasses three phases: key generation, confusion, and diffusion. The hash function and hyperchaotic are used to generate a robust key as the hash function is a one-time pad and chaotic produces unpredictable and random numbers. The suggested algorithm uses the original image and a user-defined key to generate the encryption key, thereby preventing the chosen/known-plaintext attack. In the confusion phase, there are two options to change the pixel's locations, either changing the location on the image's level or the block's level based on the Duffing map. After that, each pixel is subjected to two consecutive confusion steps: Henon and Arnold map. In the diffusion phase, the confusion phase output undergoes two successive diffusion steps: DNA followed by RSA cryptography. Using two steps in each phase maximizes the security and unpredictability of the suggested approach. Moreover, the suggested approach can withstand different attacks. Various security tests demonstrate the approach's effectiveness in withstanding attacks and achieving low correlation between neighboring pixels.

In future research, we will explore multi-model image encryption by combining color, texture, and depth data. Additionally, we aim to integrate machine learning techniques to optimize encryption parameters and enhance security. Another priority is reducing encryption time.

Use of AI tools declaration: The authors have not used Artificial Intelligence (AI) tools in creating this article.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

### REFERENCES

[1] J. Wang and W. Jiang and H. Xu and X. Wu and J. Kim, "Image encryption based on Logistic-Sine self-embedding chaotic sequence", *Optik*, vol. 271, p. 170075, 2022.

[2] N. Rani, V. Mishra and S.R. Sharma, "Image encryption model based on novel magic square with differential encoding and chaotic map", *Nonlinear Dynamics*, vol. 110, No. 3, 2022.

[3] C. Tu, R. Cui, and K. Liu, "Design of Clothing with Encrypted Information of Lost Children Information Based on Chaotic System and DNA Theory", *Autex Research Journal*, vol.0, no.0, 2022,

[4] X. Zhang and Y. Hu, "Multiple-image encryption algorithm based on the 3D scrambling model and dynamic DNA coding", *Optics & Laser Technology*, vol. 141, p. 107073, 2021.

[5] H. Hu, Y. Cao, J. Xu, C. Ma and H. Yan, "An Image Compression and Encryption Algorithm Based on the Fractional-Order Simplest Chaotic Circuit," in *IEEE Access*, vol. 9, pp. 22141-22155, 2021.

[6] X. Wang, Y.Su, L. Liu, H. Zhang, and S. Di, "Color image encryption algorithm based on Fisher-Yates scrambling and DNA subsequence operation", *Visual Computer*, vol. 47, no. 10, 2021.

[7] Y. Xiao and Z.R. Lin and Q. Xu and J. Du and L.H. Gong, "Image encryption algorithm based on semi-tensor product theory", *Journal of Modern Optics*, vol. 69, no. 19, pp. 1063-1078, 2022.

[8] X. Wang and R. Si, "A new chaotic image encryption scheme based on dynamic L-shaped scrambling and combined map diffusion", *Optik*, vol. 245, p. 167658, 2021.

[9] N. Ying, Z.Xuncai, "An Image Encryption Algorithm Based on Filling Curve and Adjacent Pixel Bit Scrambling", *Journal of Electronics & Information Technology*, vol. 44, no. 3, pp. 1137-1146, 2022.

[10] S. Wang, Q. Peng, and B. Du, "Chaotic color image encryption based on 4D chaotic maps and DNA sequence", *Optics and Laser Technology*, vol. 148, p. 107753, 2022.

[11] J. Yu, W. Xie, Z. Zhong, and H. Wang, "Image encryption algorithm based on hyperchaotic system and a new DNA sequence operation", *Chaos, Solitons and Fractals*, vol. 162, p. 112456, 2022.

[12] C. Zou, X. Wang, C. Zhou, S. Xu and C. Huang, "A novel image encryption algorithm based on DNA strand exchange and diffusion", *Applied Mathematics and Computation*, vol. 430, p. 127291, 2022.

[13] B. Jasra and A. Moon, "Color image encryption and authentication using dynamic DNA encoding and hyperchaotic system", *Expert Systems with Applications*, vol. 206, p. 117861, 2022.

[14] X. Li, J. Zeng, Q. Ding, and C. Fan, "A Novel Color Image Encryption Algorithm Based on 5-D Hyperchaotic System and DNA Sequence", *Entropy*, vol.24, no. 9, p. 1270, 2022.

[15] N. Rani, S. Rani Sharma, and V. Mishra, "Grayscale and colored image encryption model using a novel fused magic cube", *Nonlinear Dynamics*, vol. 108, pp. 1773-1796, 2022.

[16] J. Zheng and Q. Zeng, "An image encryption algorithm using a dynamic S-box and chaotic maps", *Applied Intelligence*, vol. 52, pp. 15703-15717, 2022.

[17] N. Parekh, L. D'Mello, "CHaDRaL: RGB image Encryption based on 3D Chaotic Map, DNA, RSA and LSB", 2021 International Conference on Artificial intelligence and machine vision (AIMV), 2021.

[18] M. Liu and G. Ye, "A new DNA coding and hyperchaotic system based asymmetric image encryption algorithm", *Mathematical Biosciences and Engineering*, vol. 18, pp 3887-3906, 2021.

[19] U.H. Mir, D. Singh, and P.N. Lone, "Color image encryption using RSA cryptosystem with a chaotic map in Hartley domain", *InformationSecurity Journal: A Global Perspective*, vol. 31, issue 1, pp. 49-61, 2022.

[20] K. Jiao, G. Ye, Y. Dong, X. Huang, and J. He, "Image encryption scheme Based on a Generalized Arnold map and RSA Algorithm", *Security and Communication Networks*, vol. 2020, pp. 1-14, 2020.

[21] Babu M, G. Shamala Devi, M. Yamini Krishna, M. Viswa Prasanna, N. Iswarya, "Image Encryption Using Chaotic Maps and DNA Encoding", *Journal of Xidian University*, vol. 14, issue 4, pp 1817-1827, 2020.

[22] M. Kumar, A.Saxena, S.S.Vuppala, "A survey on chaos-based image encryption techniques" " *Multimedia security using chaotic Maps: principles and Methodologies*, vol. 884, pp. 1-26, 2020.

[23] K.C.Jithin, Syam Sankar, "Colour image encryption algorithm combining Arnold map, DNA sequence operation, and a Mandelbrot set", *Journal of Information Security and Applications*, vol. 50, pp 1-22, 2020

[24] A. Akhshani, A. Akhavan, A.Mobaraki, S.C. Lim, and Z. Hassan, "Pseudo random number generator based on quantum chaotic map", *Commun Nonlinear Sci NumerSimulat* 19, vol. 19, issue 1, pp 101-111, 2014.

- [25] S.A. Elsaid, E.R. Alotaibi, and S. Alsaleh, "A robust hybrid cryptosystem based on DNA and Hyperchaotic for images encryption", *Multimedia Tools and Applications*, vol. 82, pp 1995-2019, 2022.
- [26] S.F. Yousif, A.J. Abboud, R. S. Alhumaima, "A new image encryption based on bit replacing, chaos and DNA coding techniques", *Multimedia Tools and Applications*, vol. 81, pp 27453-27493, 2022.
- [27] S. Mansoor, P. Sarosh, S.A. Parah, Habib Ullah, Mohammad Hijji, and Khan Mouhammad, "Adaptive Color Image Encryption Scheme Based on Multiple Distinct Chaotic Maps and DNA Computing", *Mathematics*, vol 10, 2022.
- [28] R.W. Ibrahim, H. Natiq, A.AlKhayyat, A.K. Farhan, N. MG. AlSaidi, D.Baleanu, "Image Encryption Algorithm Based on New Fractional Beta ChaoticMaps", *Computer Modeling in Engineering and Sciences*, vol. 131, pp.119-131, 2022.
- [29] X. Liu, X. Tong, Z.Wang, M. Zhang, "A novel hyperchaotic encryption algorithm for color image utilizing DNA dynamic encoding and self-adapting permutation", *Multimedia Tools and Applications*, vol. 81, p. 21779, 2022
- [30] M.B. Farah, A. Farah, T. Farah, "An image encryption scheme based on a new hybrid chaotic map and optimized substitution box", *Nonlinear Dynamic*, vol. 99, pp 3041–3064, 2020.
- [31] N. Iqbal, R. Naqvi, M. Atif, M.A. Khan, M. Hanif, S. Abbas, and D. Hussain, "On the Image Encryption Algorithm Based on the Chaotic System, DNA Encoding, and Castle," in *IEEE Access*, vol. 9, p. 118253, 2021.

# Object Detection and Recognition in Remote Sensing Images by Employing a Hybrid Generative Adversarial Networks and Convolutional Neural Networks

Dr Araddhana Arvind Deshmukh<sup>1</sup>, Mamta Kumari<sup>2</sup>, Dr. V.V. Jaya Rama Krishnaiah<sup>3</sup>, Suraj Bandhekar<sup>4</sup>, R. Dharani<sup>5</sup>

Head and Associate Professor, Department of Artificial Intelligence and Data Science,

Marathwada Mitra Mandal College of Engineering-Affiliated to Savitribai Phule Pune University<sup>1</sup>

Assistant professor, Department of CSE-ET, Panipat Institute of Engineering and Technology (PIET), Samalakha<sup>2</sup>

Associate Professor, Department of Computer Science and Engineering,

Koneru Lakshmaiah Education Foundation, Vaddeswaram, India<sup>3</sup>

Reader, Department of Mechanical Engineering, Rungta College of Engineering and Technology, Bhilai, C.G, India<sup>4</sup>

Assoc. Prof, IT, Panimalar Engineering College Chennai, India, 600123<sup>5</sup>

**Abstract**—Due to diverse backdrops, scale fluctuations, and a lack of annotated training data, the identification and recognition of objects in remote sensing images present major problems. In order to overcome these difficulties, this work suggests a novel hybrid technique that blends GAN and CNN. The suggested approach expands the small labelled dataset by synthesising realistic training examples using the generative abilities of GANs. The samples generated capture the various variances and backgrounds found in remote sensing photos, improving the object identification and recognition model's capacity to generalise. Additionally, CNNs, which are recognised for their outstanding feature extraction skills, are incorporated into the hybrid approach, enabling precise and reliable object identification and recognition. The model's CNN component is developed using both real and synthetic data, effectively combining the advantages of both fields. Several experiments are conducted on a large dataset of satellite photos to evaluate the performance of the proposed method. The results demonstrate that the hybrid model, with accuracy 97.32%, outperforms traditional approaches and pure CNN-based approaches in terms of dependability and resilience. The model may be efficiently generalised to unknown remote sensing images thanks to the GAN-generated samples, which bridge the gap among synthetic and actual data. The hybrid methodology used in this study demonstrates the possibility of merging GANs and CNNs for item detection and recognition using deep learning in remote sensing images.

**Keywords**—Object detection; Generative Adversarial Networks (GAN); Convolutional Neural Networks (CNN); deep learning; remote sensing; satellite images; hybrid model

## I. INTRODUCTION

Remote sensing images captured by satellites and aerial platforms provide a wealth of valuable information about the Earth's surface. Analysing these images for object detection and recognition tasks is of utmost importance in various domains such as environmental monitoring, urban planning, and disaster management [1]. Deep learning algorithms have

the potential to significantly increase the precision and effectiveness of item recognition and detection in this field when applied to remote sensing photos. Conventional methods to identifying and recognising objects in remote sensing photos frequently depend on rule-based algorithms and hand-crafted features, which have difficulties capturing the intricate and varied aspects of the data. The identification and classification of objects based upon their visual patterns and properties is made possible by deep learning techniques, which excel at autonomously learning hierarchical representations straight from the data [2].

In recent years, deep learning-based approaches have gained traction in remote sensing applications, leveraging the power of CNNs to learn discriminative features from large-scale remote sensing datasets. These models can effectively detect and recognize various objects, such as buildings, roads, vehicles, vegetation, and water bodies, in remote sensing images. By learning from a vast amount of data, CNNs can capture intricate spatial and spectral information, enabling accurate and robust object detection and recognition [3]. The advantages of deep learning in remote sensing imagery include its ability to handle complex scenes with diverse backgrounds, variations in lighting conditions, and different sensor characteristics. Additionally, deep learning models can learn from a wide range of remote sensing data sources, including optical imagery images and multispectral/hyperspectral data, making them versatile for different remote sensing applications [4]. By employing deep learning techniques, one can anticipate significant improvements in object detection and recognition performance in remote sensing images. The automated and efficient nature of deep learning models will enable faster analysis of large-scale datasets, leading to timely and accurate decision-making in various domains [5].

Additionally, the adaptability of deep learning approaches allows for transfer learning, where models trained on one remote sensing dataset can be fine-tuned on another dataset,

reducing the need for extensive annotation efforts. The effectiveness of the deep learning-based object identification and recognition system will be assessed throughout this research using benchmark satellite imagery datasets, comparing it with current state-of-the-art techniques [6]. One will consider metrics such as detection precision, recall, and computational efficiency to assess the accuracy and efficiency of our proposed approach. By advancing the state-of-the-art in deep learning-based object detection and recognition in remote sensing images, this research has the potential to greatly enhance our understanding of the Earth's surface and enable informed decision-making in a wide range of applications. The accurate identification and classification of objects in remote sensing images contribute to improved land cover mapping, infrastructure monitoring, disaster response, and environmental assessments, ultimately leading to more effective and sustainable management of our planet's resources [7].

A fundamental aspect of a computer vision job, object detection has a wide range of uses in automation, autonomous vehicles, and surveillance. By extracting discriminative characteristics from big datasets, deep learning models in particular CNN have achieved extraordinary performance in object recognition over time [8]. However, traditional CNN-based approaches often struggle with detecting objects in challenging scenarios, such as occlusions, small object sizes, and cluttered backgrounds. To address these challenges and improve object detection performance, a hybrid approach that combines the power of GANs and CNNs has gained significant attention. Generative Adversarial Networks have demonstrated their effectiveness in generating realistic synthetic data that follows the same distribution as the real data. GANs consist of a generator network and a discriminator network that engage in a competitive learning process [9]. The generator network synthesizes samples, aiming to fool the discriminator into classifying them as real, while the discriminator network tries to accurately distinguish between real and synthetic samples. This adversarial training leads to the generation of synthetic data that closely resembles the real data distribution [10].

By leveraging the generative capabilities of GANs, the hybrid approach aims to improve object detection performance by generating additional training samples. These synthetic samples provide the CNN-based object detection model with a more diverse and comprehensive understanding of object classes, augmenting the training data and enhancing the model's ability to generalize to different variations and challenging scenarios [11]. The hybrid approach involves two main stages. In the first stage, a GAN is trained on a large dataset of real object images, learning the underlying data distribution and generating synthetic samples that closely resemble real objects. These synthetic samples, combined with the real training data, create an augmented dataset for training the CNN-based object detection model. In the second stage, the CNN learns discriminative features from the augmented dataset, enabling accurate and robust object detection [12].

Fig. 1 represents the hybrid approach which offers several advantages in object detection. Firstly, it addresses the challenge of limited training data by synthesizing additional samples that capture a broader range of object variations. This augmentation leads to improved generalization and better handling of rare or underrepresented object classes. Secondly, the adversarial training process in GANs encourages the generation of realistic and diverse synthetic samples, effectively enhancing the model's ability to handle variations in object appearance, scale, and background clutter. Lastly, the hybrid approach promotes the transferability of learned features across different datasets and domains, enabling the model to adapt and generalize well to unseen data [13]. This approach focuses on developing and evaluating the hybrid approach of GANs and CNNs for object detection. This work will conduct extensive experiments using benchmark object detection datasets, comparing the performance of the hybrid approach against traditional CNN-based methods. This work will evaluate metrics such as detection accuracy, precision, recall, and robustness to challenging scenarios to assess the effectiveness of the hybrid approach [14].

CNN and GAN have revolutionized the field of object detection by providing powerful tools for accurate and robust identification of objects in images and videos. CNNs are extensively used in the early stages of object detection to extract relevant features from the input data. These deep neural networks are trained on large datasets to learn hierarchical representations of objects, enabling them to recognize patterns and objects at different levels of abstraction [15]. The convolutional layers of CNNs perform local feature extraction, while the fully connected layers analyze the extracted features and classify the objects. On the other hand, GANs play a crucial role in enhancing object detection by generating realistic and high-quality synthetic data. By training a GAN on a large dataset, it learns to generate images that closely resemble real-world objects, even in complex scenarios or rare situations. These synthetic images can be combined with the original dataset to augment the training data, thus increasing the diversity and robustness of the object detection model [16]. The improved accuracy and robustness of object detection have implications in various real-world applications, including autonomous systems, surveillance, and object recognition [17]. The findings from this research contribute to advancing the field of object detection and pave the way for more effective and reliable computer vision systems in practical applications [18]. The goal of this project is to create an effective recognition and detection of objects system for satellite or other aerial platform-derived remote sensing photos. In order to overcome issues like changing lighting circumstances and sensor noise, the project intends to automate the detection and classification of things like roads, structures, and automobiles in these photos. To increase the effectiveness of analysing remote sensing data for uses like urban planning as well as disaster assessment, a precise and scalable system is being developed.

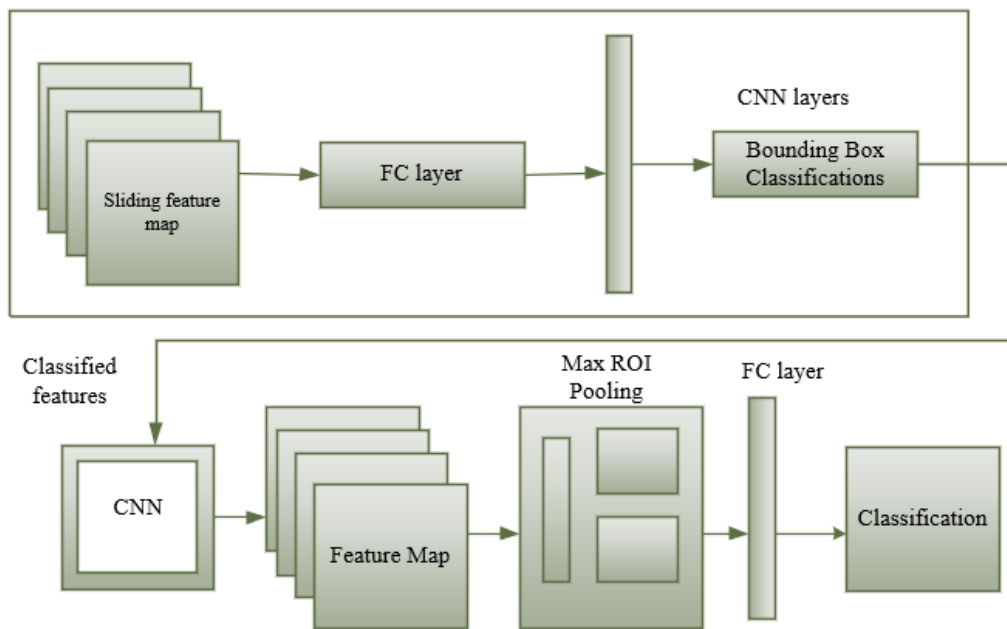


Fig. 1. CNN-GAN approach for object detection.

The following are the research's main contributions:

- The GAN-CNN approaches have been employed by other researchers, the proposed method stands out by introducing a distinctive data augmentation strategy that leverages GANs to generate authentic training examples, effectively addressing the challenges posed by diverse backgrounds, scale fluctuations, and limited annotated training data in remote sensing imagery.
- This hybrid strategy effectively augments the small annotated dataset and captures a variety of background fluctuations by harnessing the generative powers of GANs to create realistic training samples.
- The method makes the most of both domains by training the CNN components on both real and artificial data.
- The hybrid model significantly outperforms traditional and pure CNN-based approaches, according to experimental results.
- The potential of the hybrid technique for reliable and effective remote sensing item detection and recognition is demonstrated by the bridges of synthetic and actual information through GAN-generated samples, which improves the model's generalization to unknown remote sensing images.

The rest of the paper is structured as follows: Section II is described as related works while problem statement is explained in Section III. Similarly, Section IV is described as proposed methodology and Section V described as results and discussion and the conclusion is described as Section VI.

## II. RELATED WORKS

Li et al. [19] proposed a novel lightweight CorrNet is an ORSI-SOD approach. In CorrNet, first a compact subnet is

created for feature extraction and lessens the core network (VGG-16). Then initial crude prominence map is created using semantic features that are high-level in the correlation module, according to the coarse-to-fine technique. The granular scalar maps act as a geographical cue for low-level characteristics. Using the cross-layer association procedure the object position information is mined between high-level semantic characteristics. Finally, using low-level detailed characteristics, the coarse prominence map in the refinement subnet was refined to create the final fine saliency map. By lowering the requirements and calculations for each component, CorrNet ends up with only 4.09 million parameters and uses 21.09 gigaflops to execute. Results from tests on two open data sets demonstrate that lightweight CorrNet outperforms 26 modern techniques, including 16 huge methods based on CNN and two ultralight techniques, while saving a substantial amount of memory and runtime. Compact CorrNets are less suited to tackling difficult tasks involving the need for a greater capacity model because they are often built to contain less information. A lightweight CorrNet might not have enough capacity to catch such subtleties if the problem you're attempting to solve contains complex patterns or necessitates a lot of data presentation.

Sun et al. [20] created a part-based convolution neural network (PBNet) for integrated composite object detection in remote sensing pictures. PBNet evaluates an amalgamated object as a collection of parts and integrates component variables with contextual data to improve composite object recognition. Accurate ingredient knowledge can aid in the forecasting of an integrated item and help with problems resulting from various shapes and sizes. In order to provide accurate part information, a part placement module is developed that teaches the classification and localization of component positions using solely a boundary annotation. From a publicly available dataset, three representative categories of composite items are chosen for conducting

operations to test the effectiveness and generalizability of this method's identification capabilities. This dataset includes sewage treatment facilities from seven Yangtze Valley cities, encompassing an extensive variety of geographic areas. Extensive testing on two datasets demonstrates that PBNNet outperforms the current detection methods and reaches cutting-edge accuracy. Part-based models, however, primarily rely on precise part identification. The component detection process' noise or imprecision can have a negative impact on how well the PBNNet performs. Because of its reliance on precise component localization, the model may be more susceptible to mistakes or noise during the part identification process, which might result in poorer robustness and generalization.

Zhang et al. [21] proposed the Feature Pyramid Network which makes use of the built-in multiple scales rounded characteristics as well as incorporates the strong-semantic, and the weak-semantic, excellent quality features simultaneously, has been proposed as an efficient region-based VHR remote sensing imagery identification framework. The DM-FPN is made up of two modules that may be trained end-to-end: a multi-scale region suggestion network and a multiple habitats object detection network. To broaden the range of training data and get beyond input image size limitations, a number of multi-scale training methodologies are presented. To improve detection performance, particularly for tiny and dense objects, multi-scale prediction techniques are presented. Extensive tests and thorough analyse on a sizable DOTA dataset show how successful the suggested architecture. DM-FPN introduces an additional level of complexity compared to the original FPN. The inclusion of double multi-scale features requires more computational resources, including memory and processing power. This increased complexity can impact training and inference times, making it less suitable for real-time or resource-constrained applications.

Chen et al. [22] proposed a CNN for object recognition that combines scene-contextual data. The environment-contextual feature pyramidal network (SCFPN), in particular, seeks to improve the bond among the objective and the scene and address issues brought on by fluctuations in target size. The network is created by repeating an accumulated remnant block in order to enhance the ability to perform extracting features. With the help of this block, the receptive field may harvest targets' deeper information and perform very well in terms of tiny object recognition. Additionally, group normalization, which separates each channel into group and determines the variance and mean for normalization within each group, is utilized to overcome the batch normalization's limitation and enhance the efficacy of the suggested model. A tough public dataset is used to validate the suggested approach. The experimental findings show that our suggested approach outperforms existing cutting-edge object identification techniques. To include scene-level contextual data, SCFPN adds further layers and calculations. This could result in higher computing demands for both inference and training. SCFPN may be less appropriate for actual time or limited in resources applications as a result of the extra complexity.

Yan et al. [23] Developed the full-scale object detection network (FSOD-Net), which is comprised of a suggested multiscale enrichment network backbone transmitted with scale-invariant regression layers (SIRLs), is a one-stage scanner. First, by integrating the Laplace kernels with less concurrent multiscale layers of convolution, MSE-Net offers the multiscale characterization improvement. Second, because SIRLs have three distinct independent extrapolation branch layers (small, medium, and large scales), full-scale object information is covered by the default discrete scale bounding boxes (bboxes) in the regression technique. A further approach employs an oval-specific scale joint loss that combines a strong L1 norm restriction with the soft max function for each regression branch layer. It can also hasten convergence and boost anticipated b-box classification results. The findings of extensive research conducted on challenging sets of data over identifying objects in aerial images (DOTA) and object identification in visual imagery from remote sensing (DIOR), which include numerous examples from various imaging platforms, show that FSOD-Net is capable of performing better than other cutting-edge one-stage detectors. FSOD-Net, tend to have a higher computational complexity compared to simpler tasks like image classification. Object detection involves not only classifying objects but also accurately localizing their positions and generating bounding box predictions. This increased complexity can result in longer training and inference times and require more computational resources.

Ming et al. [24] proposed a Critical Feature Capturing Network (CFCNet) enhancing the accuracy of detection by focusing on three areas: developing robust visualization of features, fine-tuning pre-anchored patterns, and label assignment optimization. For instance, while constructing robust key features specific to a task, researchers first isolate the classification and recurrence elements using the Polarization Attention Module (PAM). The Rotation Anchor Refinement Module (R-ARM) performs localization improvement on preset perpendicular anchors to create superior rotation anchors using the retrieved selective regression characteristics. After that, high-quality anchors are adaptively chosen using the Dynamic Anchor Learning (DAL) technique based on their capacity to capture crucial information. The proposed system achieves outstanding performance immediate time object recognition and more potent conceptual representations for structures in remote sensing pictures. Experimental finding on three remote sensing datasets which demonstrate that this technique outperforms numerous state-of-the-art methodologies in terms of detection performance. Attention mechanisms often require additional memory to store the attention maps or weights. If PAM generates attention maps with high spatial resolution, it can significantly increase the memory usage of the network, making it less suitable for memory-constrained environments or large-scale applications.

Lu et al. [25] proposed a feature-fusion SSD and an end-to-end network called attention. First, a complex feature fusion framework is created to improve the shallow features' semantic information. The feature information is then screened by the introduction of a dual-path attention module. The background noise is muted and the main feature is

highlighted in this module using spatial focus and channel attention. A multiscale responsive field module follows, which improves the network's capacity for feature representation even more. In order to correct the imbalance among both the positive and negative samples, the loss function is lastly optimized. The results of this method's experiments on the data sets demonstrate its efficacy. Integrating attention mechanisms and feature fusion techniques into the SSD framework can introduce additional computational overhead. This may lead to increased training and inference times, making it less suitable for real-time or resource-constrained applications. SSD is designed to handle objects of different scales using a set of predefined anchor boxes. The introduction of attention and feature fusion mechanisms may introduce additional challenges in effectively handling scale variation. If not properly designed, the architecture may struggle to accurately detect objects at various scales, leading to potential detection errors. Attention mechanisms and feature fusion techniques introduce additional learnable parameters into the model. This increased parameter count may make the model more prone to overfitting, especially when the training dataset is limited or regularization techniques are not effectively employed. Careful parameter initialization and regularization strategies are required to mitigate these issues.

Fu et al. [26] suggested a feature-fusion architecture that uses a top-down pathway to add semantic descriptions to depth layer characteristics and an upward pathway to integrate top layer map features with low-level data to produce a multiple scales feature hierarchy. It is possible to create a potent representation of characteristics for numerous scales objects by mixing features from many levels. Axis-aligned boxes, which may include nearby instances and backdrop regions, have been used by the majority of prior approaches to find objects with variable directions and dense spatial distributions. This method creates a rotation-aware entity detector that locates items in remote sensing pictures by using oriented boxes. The region suggestion network adds more default angles to the anchors to better cover orientated objects. Instead of using horizontal proposals, which only imperfectly locate oriented objects, it uses oriented suggestion boxes to contain objects. For obtaining the characteristic maps of oriented suggestions for the next R-CNN subnetwork, the orientation-based RoI pooling procedure is implemented. On a public dataset, extensive tests are run for oriented object recognition in remote sensing photos. Feature-fusion architectures typically rely on having access to multiple modalities or sources of information. If one or more of these modalities are missing or inaccessible, the model may not be able to effectively perform feature fusion, limiting its performance or applicability.

### III. PROBLEM STATEMENT

The problem addressed in this research is the development of a robust and efficient object detection and recognition system for remote sensing images. Remote sensing images, acquired from satellites or aerial platforms, provide valuable information for applications such as land cover mappings, urban planning, and disaster assessment. However, manually

analysing these images is time-consuming and impractical, necessitating automated methods to identify and classify objects of interest [27]. The objective of this study is to design an accurate and scalable system that can detect and localize various objects in remote sensing imagery, such as buildings, roads, vehicles, and natural features, and subsequently recognize and categorize them into relevant classes. The system should be able to handle the challenges associated with remote sensing data, such as varying lighting conditions, sensor noise, and the large-scale nature of the datasets. By addressing these challenges, the proposed system aims to enhancing the efficiency and accuracy of object detection and recognition in remote sensing images, facilitating the analysis and interpretation of these critical data sources.

### IV. PROPOSED GAN-CNN APPROACH

For object detection and recognition in remote sensing images, the suggested methodology combines GANs with CNNs. The studies make use of two datasets, DOTA and UCAS-AOD, totalling 2900 and 1500 aerial photos with labelled items, respectively. By producing fake remote sensing photos, GAN-based data augmentation is used to broaden the dataset's variety and generalisation. An object generation network plus an image interpretation network makes up the GAN model known as RDAGAN. The image interpretation network makes sure that the generated pictures approximate the specified domain while the object generation network creates realistic objects. In order to handle multiscale objects, the CNN-based object identification employs a Faster R-CNN architecture with multilayer Region Proposal Networks (RPNs). The RPNs improve the detection of both tiny and large objects by using various CNN levels to create object suggestions. Additionally, CNN feature map fusion is included in the suggested approach to improve the representation of tiny objects without the need of up sampling. Overall, to accomplish precise and reliable object detection and recognition, the hybrid strategy includes GAN-based data augmentation, CNN-based object detection, and specialised algorithms for remote sensing images. Fig. 2 shows the Overall architecture of the proposed methodology.

#### A. Dataset Collection

DOTA and UCAS-AOD are two datasets used in the study's experiments [28]. The responsibilities for oriented (OBB) and horizontal bounding boxes (HBB) are included in both. The DOTA dataset, which now comprises 2900 aerial images with pixel sizes ranging from 800 x 800 to 4000 x 4000 and objects belonging to fifteen different groups with an overall number of 196171 occurrences, is the biggest dataset for object recognition in aerial imagery. It is divided into three sets: training (1/2), validation (1/6), and testing (1/3). UCAS-AOD includes 15683 occurrences of each of the two classifications (Plane and Car) and 1500 aerial images, each measuring roughly 1000 by 1000 pixels. For training and assessment, the research randomly chose 1220 images. Employing the authorized development kit for DOTA, the study divided images into  $1024 \times 1024$  squares with 512 pixels of overlaps. The datasets for identifying objects and recognition using remote sensing are displayed in Table I.

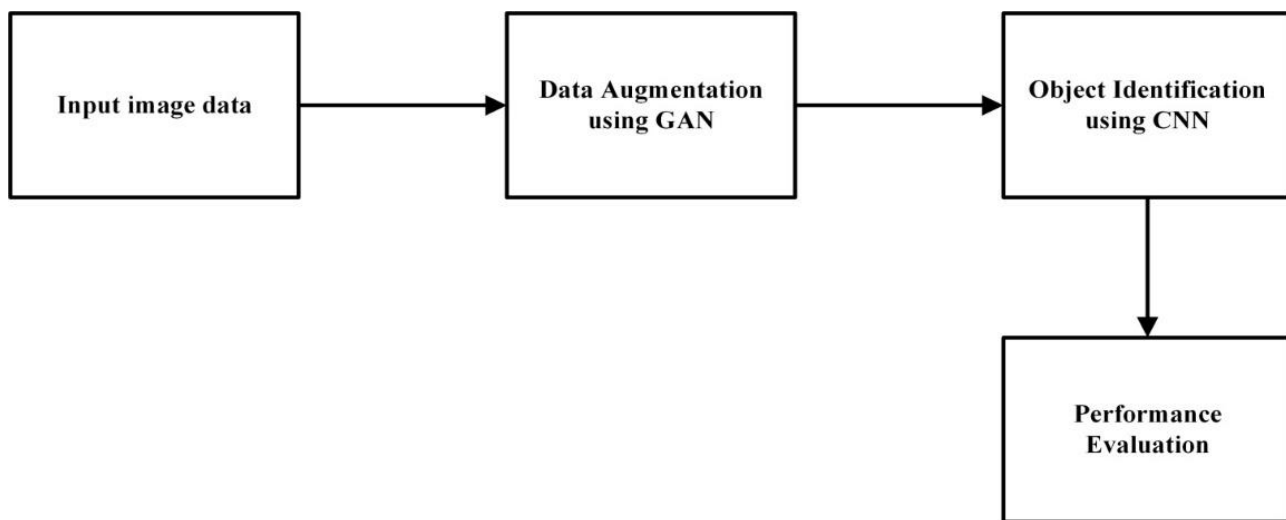


Fig. 2. Overall architecture of the proposed methodology.

TABLE I. REMOTE SENSING OBJECT DETECTION AND RECOGNITION DATASETS

Datasets	Total number of images	Categories	Instances	Image Sizes (Pixels)	Image Type
DOTA	2900	15	196171	800x800 to 4000x4000	RGB
UCAS-AOD	1500	2	15683	1000x1000	RGB

### B. GAN based Data Augmentation

In several industries, including remote sensing and medical imaging, an image data augmentation technique based on GAN is frequently employed. Because neural networks in these domains need a lot of training data, it might be challenging to collect enough of it. It is simple for models to over fit or fall victim to the class imbalance problem when there are few data points. By creating fresh samples from a data distribution, the GAN-based picture data augmentation techniques can solve these issues. Fig. 3 shows the overall design of RDAGAN.

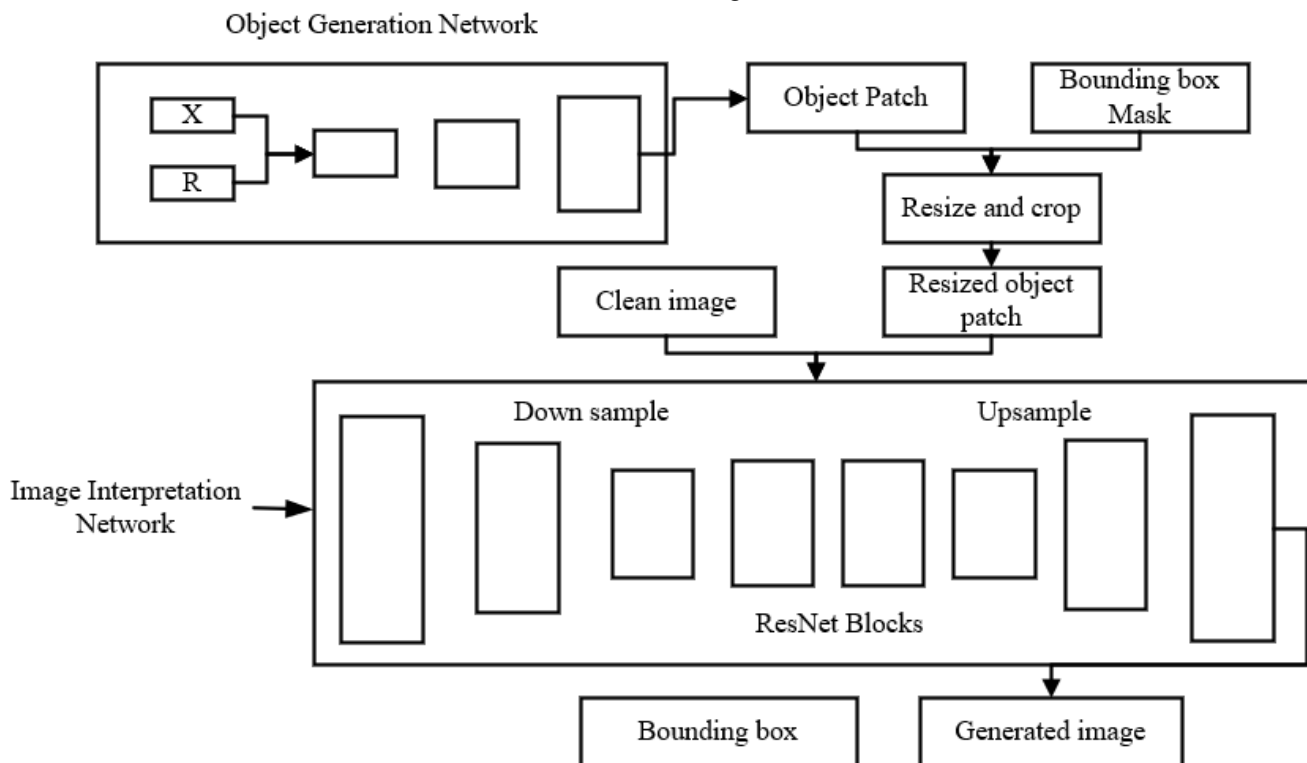


Fig. 3. The overall design of RDAGAN.



RDAGAN operates by training a GAN to generate synthetic data samples that closely resemble the true data distribution. This process significantly expands the available training data, mitigating the risk of over fitting and enhancing model generalization. The suggested robust data augmentation GAN (RDAGAN) model does the data augmentation. The objective was to create a framework that maps targeted images in the desired image domain ( $i_s \in S$ ) to cleaned images in the image cleanliness domain ( $i_r \in R$ ). The proposed algorithm was trained employing a dataset for object recognition with few images, most of which contained occlusions. The proposed framework uses the divide-and-conquer strategy, splitting the framework into two networks. The model tries to add realistic objects to the image ( $i_r$ ), but it also tries to change the overall image to seem like it belongs in the domain that it is targeting ( $S$ ). A single GAN structure makes it difficult to accomplish these objectives since the training process becomes unpredictable.

To be placed into  $i_r$ , the target object's image is created by the object creation network. The visualization created by the object creation network is sent into the image interpretation network as a source of input. Due to the objectives of object formation and image interpreting, the imagery reduces the training instabilities in the image interpretation network. In order to generate a separated illustration of the target object, the network utilizes the InfoGAN architecture. The image interpretation network creates a loss function using the separated representations it received from the object development network. The crop and resize modules  $C$  were used to crop and resize object images  $C(i_c)$  from image  $i_r$  in order to train the algorithm. The hidden code  $r$  and indestructible noise  $x$ , which are obtained by sampling from a normal distribution, are inputs to the generator  $E_{obj}$ . In addition to validating the input images, the discriminator  $M_{obj}$  also forecasts the input hidden code  $r'$ .

The framework's goal  $N_{obj}$  includes two losses since the object formation network utilizes the InfoGAN design: an adversarial loss  $N_{GAN}^{obj}$  and an information loss  $N_{Info}^{obj}$ .

The negative outcome Eq. (1) explains how to utilize  $N_{GAN}^{obj}$  to ensure that the created patch  $E_{obj}(x, r)$  resemble the domain of the intended images  $C(i_s)$ .

$$N_{GAN}^{obj} = G_{C(i_s) \sim S} \log M_{obj}(C(i_s)) + G_{C(i_r) \sim R} \log(1 - M_{obj}(E_{obj}(x, r))) \quad (1)$$

The interaction of data between the produced image  $G_{C(i_r)}$  and the hidden code  $c$  is measured by loss of data  $N_{Info}^{obj}$ . Eq. (2) explains how to compute it employing the mean squared error of the projected code  $r'$  from the discriminator  $M_{obj}$  and the input hidden code  $c$ .

$$N_{Info}^{obj} = G_{r \sim L(0,1), r' \sim M_{obj}(E_{obj}(x, r))} (\|r - r'\|^2) \quad (2)$$

Eq. (3) states that the total goal,  $N_{obj}$ , is the aggregate of prior losses.

$$N_{obj} = N_{GAN}^{obj} + \lambda N_{Info}^{obj} \quad (3)$$

Where the extent of the data loss is represented by minimizing the overall aim, a simulation was trained.

The intended image  $i_s \in S$  is created by combining the freshly processed images  $i_r \in R$  and the object patch  $E_{obj}(x, r)$  produced by the object patch network using the image interpretation network  $i_r$ . However, utilizing the standard GAN model and a single adversarial loss, it is difficult to carry out these difficult tasks at once. To lessen the load of difficult jobs, the suggested model contains a local discriminator and extra loss functions.

1) *Generator*: Similar to the generator employed in CycleGAN, the image interpretation network generator  $E_{sc}$  features encoder-decoder architecture with blocks from the residual network (ResNet) in the center. However, because every characteristic are down and up sampled, the generator has adaptability in the form variance of the output imagery.

The generator needs a bounding box mask  $d_a$ , which identifies the place of flame insertion, and to produce the image. Eq. (4) demonstrate that the place where the mask's value is 0 denotes the background and the position where its value is 1 denotes the flame's location. The bounding box region is determined using no special techniques. The height and breadth of the images are used to sample discrete uniform randomness at each location in the bounding box region.

$$d_a = \begin{cases} 1 & \text{for flame} \\ 0 & \text{for background} \end{cases} \quad (4)$$

By resizing the object patch and placing it in the region where the integer value of the bounding box mask is one, the resized object patch  $i_q := \text{Resize}(E_{obj}(x, r))$  is created. The generator input is created by concatenating the scaled object patch with a clean imagery. By automatically combining the six-channel combined imagery and interpreting them such that they resemble the intended domain image  $i_s \in S$ , the generator produces the produced image  $E_{sc}(i_q, i_r)$ .

2) *Discriminator*: The global  $M_{sc}^{global}$  and the local  $M_{sc}^{local}$  discriminators make up the image interpretation network. The image interpretation network responsibilities of image interpretation and natural merging are carried out by these discriminators.

The images produced by the generator,  $E_{sc}(i_q, i_r)$ , are evaluated by the global discriminator,  $M_{sc}^{global}$ . The PatchGAN discriminator, which analyses portions of the image rather than the entire one, serves as the foundation for its construction. It determines if the imagery is comparable to the intended domain image  $S$ . An adversarial loss results from this assessment outcome.

When utilizing the mask of the created image  $E_{sc}(i_q, i_r)$ , the local discriminator  $M_{sc}^{local}$  decides if the object patch  $C(E_{sc}(i_q, i_r))$  is realistic and whether it can be acquired through the scaling and cropping operation  $R$ . The local discriminator's architecture is comparable to that of the global discriminators. However, similar to the InfoGAN discriminator, it also has a separate auxiliary layer that

generates the anticipated code  $r'$  from the image's map of characteristics. The adversarial loss contains the local discriminator's authentic assessment outcome.

3) *Adversarial loss*: In order to illustrate the generator for the mapping from R to S, the study employed adversarial loss  $M_{GAN}^c$ . Eq. (5) represents the goal as follows:

$$\begin{aligned} M_{GAN}^{sc} = & \\ & G_{i_s \sim q_s} \log M_{sc}^{global}(i_s) + \\ & G_{i_q \sim E_{gen}(x,r), i_r \sim Q_r} \log M_{sc}^{local}(E_{sc}(i_q, i_r)) \\ & + G_{i_s \sim q_s} \log(1 - M_{sc}^{global}(C(i_s))) + G_{i_q \sim E_{gen}(x,r), i_r \sim Q_r} \\ & \log(1 - M_{sc}^{local}(C(E_{sc}(i_q, i_r)))) \end{aligned} \quad (5)$$

Where the global discriminant  $M_{sc}^{global}$  seeks to separate the produced image  $E_{sc}(i_q, i_r)$  from the images acquired from the intended domain S, whereas  $E_{sc}$  attempts to produce images identical to those received from the targeted domain S and object targets look as genuine objects. In order to distinguish the created object  $C(E_{sc}(i_q, i_r))$  from the object acquired from S, the local discriminator  $M_{sc}^{local}$  makes a determination.

### C. CNN-Based Object Detection

The foundation for CNN-based object detection is introduced in this section. In contrast to categorization, the problem of object detection requires the prediction of both the precise location and labelling of numerous objects inside an image. R-CNN was initially a very effective method of object detection in the fields of computer vision. Three processes make up R-CNN: categorization, representation of characteristics obtained by CNN, and area proposal produced by selective search. Without having to calculate each ROI, Fast R-CNN can speed up object identification. In order to create fixed-dimensional characteristics from every ROI, it applies a ROI-pooling layer. The Hyper Region Proposal Network (RPN) now includes the production of object proposals because of the Faster R-CNN. Improved accuracy is achieved via faster R-CNN, which unifies object identification and recognition into a single network. It has influenced other creative and profitable item detectors for special instances.

The objective is to develop a unique detection network that can recognize both tiny and large items by utilizing the quicker R-CNN, which identifies objects employing high-level semantics. Faster R-CNN cannot be immediately implemented in remote sensing objects recognition because to recognize the distinctions between natural and remote sensing imagery. There are other optimization techniques suggested, such as multilayer RPNs and detecting subnetworks. The characteristic representation of the image is recovered using a sequence of convolution layers in the quicker R-CNN framework. RPN employs a number of anchors with predetermined sizes and aspect ratios over the map of characteristics to generate object proposals. Convolutional characteristics along with object suggestions are used in the categorization step to determine the labelling's

and bounding box of various objects. Due to the distinctions among natural and remote sensing images, it is difficult to recognize certain tiny objects in large remote sensing images, such as vehicles and ships, and it is also important for balancing these multiscale objects because certain large objects, such as ground track fields, must be identified. The CNN built on quicker R-CNN along fails to operate well on remote sensing information in regard to all the difficulties.

1) *Multilayer RPNs*: In the attempts, the bases are raised first taking into account the RPN principles. The original CNN bases employ three ratios of aspect and three scales,  $\{128^2, 256^2, 512^2\}$ . Extremely small bases have been included to the collection of bases because small objects can be seen in remote sensing images. The study finally uses five scales  $\{32^2, 64^2, 128^2, 256^2, 512^2\}$  to accurately fit the ground truth after multiple failed tries. There are now fifteen bases instead of the previous nine bases. The reliability of particular small object detection has increased as a result of this improvement. Although adding additional bases is an easy and basic technique to find more small components, the precision still cannot be improved upon. The size of the characteristic map gets smaller as the CNN advances, and typically the last layer characteristics are input into the RPN. This causes smaller components in a big image to lose information. There may be no information about this object in the characteristic map of the previous layer. The study assumes that lowering network levels have reduced receptive fields and therefore better suited for tiny item identification. On the other hand, larger objects can be detected better at higher levels.

The VGG16 model and ResNet-101 framework are the foundations of the proposed SAPNet. There are 13 convolution layers in the VGG16. The four pooling layers can split all of the convolution layers into five segments. Faster R-CNN generates proposals using the conv5\_3 layer, although it is challenging to include the characteristics of small objects. By creating a second RPN network, the study employs conv4\_3 to forecast ROIs, in contrast to earlier techniques that exclusively used conv5\_3 to create recommendations. Considering that conv5\_3 in VGG16 acquires more characteristics to obtain huge objects whereas conv4\_3 in VGG16 has additional characteristics concerning smaller objects. These two layers are suggested for adoption by multilayer RPNs.

There are two RPNs in the proposed network, as seen in Fig. 4. The first, RPN1, utilizes the conv5\_3 layer, whereas the second, RPN2, employs the conv4\_3 layer. While RPN1 concentrates on large proposals, RPN2 concentrates on modest proposals. The multilayer RPNs may provide multiscale remote sensing object suggestions through two RPN branches. When fitting huge objects in RPN1, the scale set  $\{128^2, 256^2, 512^2\}$  is used. When generating tiny object suggestions in RPN2, the scale set  $\{32^2, 64^2, 128^2\}$  is used. RPN1 utilizes the characteristics map produced by block 5 for ResNet-101, whereas RPN2 employs the characteristic map produced by block 4.

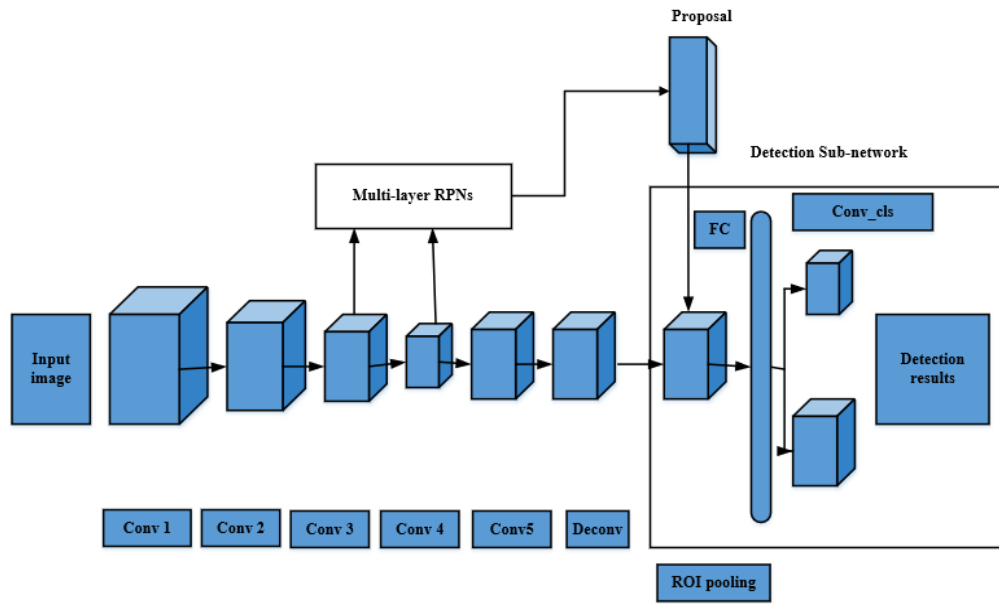


Fig. 4. Multilayered RPNs and a detection sub network.

The final identification subnetwork is where the anticipated bounding boxes and the labels of items originate from. A discrete probability,  $q = (q_0, q_1, \dots, q_k)$ , over  $K+1$  classifications ( $K$  object classes and one background class), is present in each ROI during the training phase. A ground truth label  $y$  is assigned to each ROI. In this case, the first category has been configured as the background, while classes'  $1-K$  corresponds to the classes of the ground truth. The bounding box regression target is denoted by the expression  $\hat{t} = (\hat{t}_u, \hat{t}_v, \hat{t}_x, \hat{t}_y)$ , the regressed bounding box by  $t = (t_u, t_v, t_x, t_y)$ , and the loss of every recognition layer by the expression is given in Eq. (6) and (7).

$$N(q, x, t, \hat{t}) = N_{els}(q, x) + \lambda[x \geq 1]N_{loc}(t, \hat{t}) \quad (6)$$

Where,

$$N_{loc}(t, \hat{t}) = \sum_{i \in \{u, v, x, y\}} smooth_{N_1}(t - \hat{t}) \quad (7)$$

In which,

$$smooth_{N_1(y)} = \begin{cases} 0.5y^2 \\ |y| - 0.5, & otherwise \end{cases}$$

The bounding box loss and classification loss are counterbalanced by the hyperparameter  $\lambda$ . During the test,  $\lambda$  is set to 1.

2) *CNN feature map fusion*: Certain methods simply exaggerate the input images before feeding them into the network because the pretrained CNN model only accepts input with a predetermined size ( $224 \times 224$  in VGG16, for example). These methods have an impact on the effectiveness of the detection, particularly for tiny items. Some techniques up sample the input images to correct for scale inconsistencies, but this uses more memory and slows down processing. The quicker R-CNN network's ROI pooling layer is still being studied. The network can analyse pictures of any

size thanks to its structure, which pools proposal regions into a fixed resolution of  $7 \times 7$ . Utilizing low level features is an effective approach to boost the information of tiny objects rather than up sampling the input images. The higher-level characteristics must be up sampled before being merged with the low-level characteristics since the size of the high-level characteristics is lower than that of the low-level characteristics.

## V. RESULTS AND DISCUSSION

### A. Evaluation metrics

Having into consideration the needs for practical engineering purposes, the approach was assessed using accuracy, average precision (AP), recall, frames per second (FPS), and the precision-recall (PR) curve. To clearly illustrate the results, a distinct matching rule precision-recall (PR) curve was created, and a new PR was created according to it.

1) *Precision and Recall (PR)*: PRC is a commonly employed metric utilised in numerous studies on object detection. Both the recall and accuracy measures can be written below, given that TP, FN, and FP stand for the amount of true positives, false negatives, and false positives, namely in Eq. (8)-(10)

$$\text{Precision} = TP / TP + FP \quad (8)$$

$$\text{Recall} = TP / TP + FN. \quad (9)$$

$$\text{F1 score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (10)$$

If a detection results has an intersecting over union (IOU) to a baseline value of at least 0.5, it is projected to be a true positive; alternatively, it is regarded as a false positive [29].

IoU refers to the intersection area over the union region between the two boxes with boundaries in the context of object detection. A projected boxes is considered a true

positive (TP) if the IoU corresponding to the ground-truth box ( $G_0$ ) and forecasted box ( $D_0$ ) exceeds than the standard threshold, and a false positive (FP) else. IoU is characterised by Eq. (11)

$$IoU = G_0 \cap D_0 / G_0 \cup D_0 \quad (11)$$

A ground-truth box is said to be false negative (FN) when it is unable to locate the corresponding anticipated container. One may create a PR curve using these numbers for TP, FP, and FN, accuracy, and recall following calculating dynamic recall and precision at various scoring thresholds.

**B. Average precision (AP)**

The region underneath the PR curve is known as AP. They assess the detection outcomes of the suggested strategy using the mean average precision (mAP) in Eq.. (12) [30].

$$mAP = \frac{1}{N_0} \sum_{i=1}^N AP_{0_i} \quad (12)$$

1) *Intersection-over-Detection (IoD)*: They create an additional challenging matched rule to compute TP in order to test the capacity to forecast the entire composites object. The intersection across the region of the outcome of detection is characterised by a new IoD, that is indicated by the following:

$$IoD = G_0 \cap D_0 / D_0 \quad (13)$$

IoD is more capable to show the superior performance of PBNNet than IoU. As an outcome, when  $IoD > 0.5$ , an additional PR curve (dubbed PR-IoD) can be built.

TABLE II. PERFORMANCE METRICS OF PROPOSED GAN-CNN

Metrics	Proposed Method GAN-CNN
Accuracy	97.32
Precision	96.53
Recall	94.42
F1 score	92.27

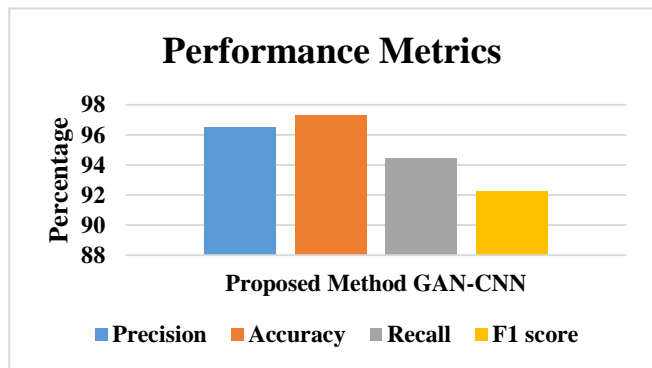


Fig. 5. Performance metrics of proposed GAN-CNN.

On the remote sensing picture dataset, the suggested GAN-CNN hybrid technique displayed outstanding results across key evaluation measures. The hybrid approach in Table II demonstrated its exceptional capacity to produce high accuracy, precision, and recall, culminating in a strong F1 score, with an accuracy of 97.32%, precision of 96.53%, recall of 94.42%, and an F1 score of 92.27% which is shown in Fig. 5. The metrics presented in Table II represent the average

performance of the proposed GAN-CNN method across both the UCAS-AOD dataset and the DOTA dataset. These findings highlight the hybrid methodology's ability to greatly increase the accuracy of object detection and recognition, giving it an attractive option for strengthening the interpretation and analysis of remote sensing data in a variety of applications.

TABLE III. PRECISION, RECALL, F1-SCORE OF EXISTING METHODS AND PROPOSED GAN-CNN [31]

Methods	Recall	Precision	F1 score	Accuracy
YOLO v3	78.09	84.62	81.22	84.86
SSD	77.35	83.36	80.24	85.94
CFF-SDN	87.23	93.11	90.07	94.68
Faster R-CNN	83.32	89.65	86.37	87.64
GAN-CNN	94.42	96.53	92.27	97.32

The YOLO v3 model demonstrated a recall of 78.09%, precision of 84.62%, F1 score of 81.22%, as well as accuracy of 84.86% in the evaluation of object recognition methods using the specified metrics on a remote sensing image dataset in Table III. Similar results were shown by the SSD model, which had an accuracy of 85.94%, a recall of 77.35%, and precision of 83.36%. With a recall of 87.23%, precision of 93.11%, F1 score of 90.07%, as well as accuracy of 94.68%, the CFF-SDN approach in particular produced better results. Recall was 83.32%, precision was 89.65%, F1 score was 86.37%, and accuracy was 87.64% for the Faster R-CNN model. The proposed GAN-CNN hybrid strategy, with recall of 94.42%, precision of 96.53%, F1 score of 92.27%, as well as accuracy of 97.32%, however, Fig. 6 demonstrated the most astounding performance across all parameters. These results highlight the hybrid approach's clear superiority over more traditional approaches, emphasising its potential to achieve extraordinarily accurate and dependable item recognition and detection in remote sensing images.

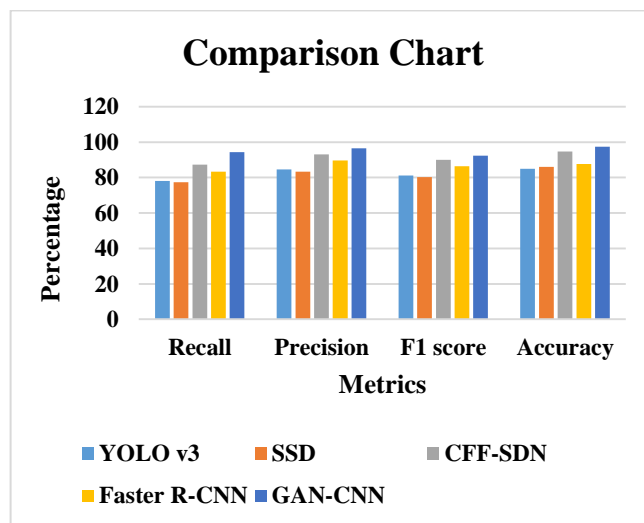


Fig. 6. Comparison chart of Precision, Recall, F1 score, Accuracy.

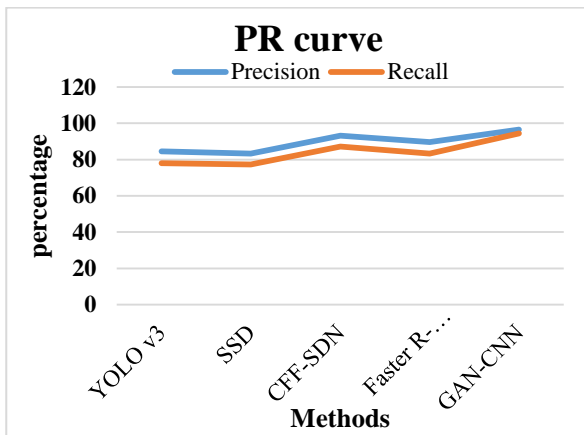


Fig. 7. PR curve of existing vs. proposed methods.

Variable performance across several approaches was demonstrated by a precision-recall curve assessment of the examined object detection algorithms on the remote sensing picture dataset in Fig. 7. While SSD demonstrated a precision of 83.36% and a recall of 77.35%, YOLO v3 attained an accuracy rate of 84.62% at a recall rate of 78.09%. The precision and recall numbers for the CFF-SDN technique were noticeably higher, with a 93.11% precision translating to an 87.23% recall percentage. Faster R-CNN achieved a recall of 83.32% and a precision rate that was 89.65%. With a precision of 96.53% and a phenomenal recall of 94.42%, the proposed GAN-CNN hybrid technique stood out as having the highest precision and recall rates. These trade-offs between high precision and recall, which are crucial for successful recognition and detection of objects activities in remote sensing images, offer insightful information about how well the models perform across various thresholds.

YOLO v3 scored a mAP of 82.73 among the investigated object detection techniques on the remote sensor image dataset in Table IV. SSD came in second with an overall rating of 81.53. With a mAP of 87.81, faster R-CNN displayed excellent performance. Notably, CFF-SDN fared better than the other approaches, obtaining a noteworthy mAP of 91.51. The maximum mAP of 94.32 was demonstrated by the suggested GAN-CNN hybrid strategy, outperforming all other approaches in Fig. 8. These findings emphasise the efficacy of the hybrid strategy and demonstrate its potential to greatly outperform both conventional single-model CNN methods and other cutting-edge methods in item recognition and detection in remote sensing images.

TABLE IV. MAP VALUES OF DIFFERENT METHODS

Methods	mAP
YOLO v3	82.73
SSD	81.53
CFF-SDN	91.51
Faster R-CNN	87.81
GAN-CNN	94.32

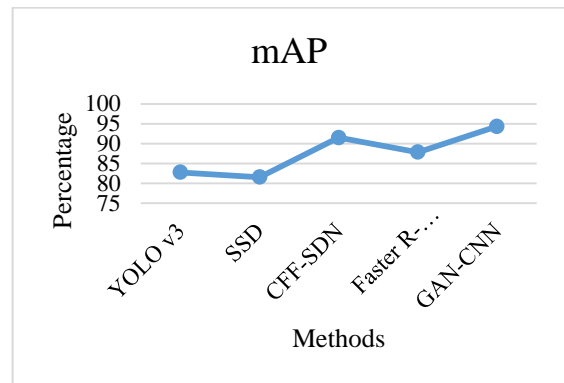


Fig. 8. Mean AP Curve for Different Methods.

## VI. CONCLUSION

The authors propose a novel hybrid GAN-CNN approach for object detection and recognition in remote sensing images, aiming to address the challenges of diverse backgrounds, scale fluctuations, and limited annotated training data. Their method stands out through a data augmentation strategy that leverages GANs to generate realistic training examples capturing remote sensing photo variations, effectively expanding the labeled dataset. Notably, they integrate both real and synthetic data into their CNN component, combining the strengths of both domains. Their approach achieves superior performance, with an accuracy of 97.32%, surpassing traditional and pure CNN-based methods, while also showcasing the ability to generalize to unknown remote sensing images, bridging the gap between synthetic and actual data and demonstrating the potential of merging GANs and CNNs for remote sensing object detection and recognition. The evaluation's findings show how this hybrid approach can improve efficiency in comparison to more conventional CNN-based techniques. The hybrid technique solves issues particular to remote sensing images, including a lack of data annotations, unbalanced class distributions, and complicated backdrops, by introducing GANs into the learning pipeline. The GAN element creates artificial examples that accurately reflect the geographic distribution of targeted objects, enhancing the variety of the information and enhancing the generalisation abilities of the CNN component. Researchers found increased object detection precision, higher identification rates, and greater adaptability to difficult backdrops through empirical assessments. The combined methodology demonstrated its supremacy in remote sensing recognition and detection of objects tests by outperforming state-of-the-art techniques. The combined technique lowers the dependency on large-scale labelled datasets, which are frequently difficult to get in the satellite imagery area, by producing artificial data using GANs. This characteristic makes the technique realistic and adaptable to real-world circumstances by enabling more effective inference and training. Although the hybrid strategy has produced encouraging results, more study is needed in several areas. The accuracy and realistic nature of samples produced might be improved by adjusting the GAN design and investigating various GAN versions. Exploring various CNN designs, hyper parameters, and training methods would also offer insightful information for enhancing the efficiency of the hybrid technique. New opportunities for effective and

precise analysis of remote sensing imagery are made possible by its capacity to handle issues unique to remote sensing data, enhance performance, and lessen the reliance on data with annotations. The use of accurate item identification and recognition in decision-making processes is crucial in many programmes, such as urban planning, agriculture, environmental monitoring, and disaster management.

#### REFERENCES

- [1] S. N. Shivappriya, M. J. P. Priyadarsini, A. Stateczny, C. Puttamadappa, and B. D. Parameshachari, "Cascade Object Detection and Remote Sensing Object Detection Method Based on Trainable Activation Function," *Remote Sensing*, vol. 13, no. 2, Art. no. 2, Jan. 2021, doi: 10.3390/rs13020200.
- [2] X. Qian, S. Lin, G. Cheng, X. Yao, H. Ren, and W. Wang, "Object Detection in Remote Sensing Images Based on Improved Bounding Box Regression and Multi-Level Features Fusion," *Remote Sensing*, vol. 12, no. 1, Art. no. 1, Jan. 2020, doi: 10.3390/rs12010143.
- [3] H. Ma, Y. Liu, Y. Ren, and J. Yu, "Detection of Collapsed Buildings in Post-Earthquake Remote Sensing Images Based on the Improved YOLOv3," *Remote Sensing*, vol. 12, no. 1, Art. no. 1, Jan. 2020, doi: 10.3390/rs12010044.
- [4] A. Mohan, A. K. Singh, B. Kumar, and R. Dwivedi, "Review on remote sensing methods for landslide detection using machine and deep learning," *Transactions on Emerging Telecommunications Technologies*, vol. 32, no. 7, p. e3998, 2021, doi: 10.1002/ett.3998.
- [5] M.-T. Pham, L. Courtrai, C. Friguier, S. Lefèvre, and A. Baussard, "YOLO-Fine: One-Stage Detector of Small Objects Under Various Backgrounds in Remote Sensing Images," *Remote Sensing*, vol. 12, no. 15, Art. no. 15, Jan. 2020, doi: 10.3390/rs12152501.
- [6] Y. Yu, J. Zhao, Q. Gong, C. Huang, G. Zheng, and J. Ma, "Real-Time Underwater Maritime Object Detection in Side-Scan Sonar Images Based on Transformer-YOLOv5," *Remote Sensing*, vol. 13, no. 18, Art. no. 18, Jan. 2021, doi: 10.3390/rs13183555.
- [7] J. Rabbi, N. Ray, M. Schubert, S. Chowdhury, and D. Chao, "Small-Object Detection in Remote Sensing Images with End-to-End Edge-Enhanced GAN and Object Detector Network," *Remote Sensing*, vol. 12, no. 9, Art. no. 9, Jan. 2020, doi: 10.3390/rs12091432.
- [8] S. S. Ismail, R. F. Mansour, A. El-Aziz, M. Rasha, A. I. Taloba, and others, "Efficient E-mail spam detection strategy using genetic decision tree processing with NLP features," *Computational Intelligence and Neuroscience*, vol. 2022, 2022.
- [9] Y. Li, Q. Huang, X. Pei, L. Jiao, and R. Shang, "RADet: Refine Feature Pyramid Network and Multi-Layer Attention Network for Arbitrary-Oriented Object Detection of Remote Sensing Images," *Remote Sensing*, vol. 12, no. 3, Art. no. 3, Jan. 2020, doi: 10.3390/rs12030389.
- [10] T. Hoese and C. Kuenzer, "Object Detection and Image Segmentation with Deep Learning on Earth Observation Data: A Review-Part I: Evolution and Recent Trends," *Remote Sensing*, vol. 12, no. 10, Art. no. 10, Jan. 2020, doi: 10.3390/rs12101667.
- [11] K. Ravikumar, P. Chiranjeevi, N. M. Devarajan, C. Kaur, and A. I. Taloba, "Challenges in internet of things towards the security using deep learning techniques," *Measurement: Sensors*, vol. 24, p. 100473, 2022.
- [12] H. Chen and Z. Shi, "A Spatial-Temporal Attention-Based Method and a New Dataset for Remote Sensing Image Change Detection," *Remote Sensing*, vol. 12, no. 10, Art. no. 10, Jan. 2020, doi: 10.3390/rs12101662.
- [13] K. Lambers, W. B. Verschoof-van der Vaart, and Q. P. J. Bourgeois, "Integrating Remote Sensing, Machine Learning, and Citizen Science in Dutch Archaeological Prospection," *Remote Sensing*, vol. 11, no. 7, Art. no. 7, Jan. 2019, doi: 10.3390/rs11070794.
- [14] U. Alganci, M. Soydas, and E. Sertel, "Comparative Research on Deep Learning Approaches for Airplane Detection from Very High-Resolution Satellite Images," *Remote Sensing*, vol. 12, no. 3, Art. no. 3, Jan. 2020, doi: 10.3390/rs12030458.
- [15] W. Li, H. Liu, Y. Wang, Z. Li, Y. Jia, and G. Gui, "Deep Learning-Based Classification Methods for Remote Sensing Images in Urban Built-Up Areas," *IEEE Access*, vol. 7, pp. 36274–36284, 2019, doi: 10.1109/ACCESS.2019.2903127.
- [16] N. Omer, A. H. Samak, A. I. Taloba, and R. M. Abd El-Aziz, "A novel optimized probabilistic neural network approach for intrusion detection and categorization," *Alexandria Engineering Journal*, vol. 72, pp. 351–361, 2023.
- [17] H. Su et al., "HQ-ISNet: High-Quality Instance Segmentation for Remote Sensing Imagery," *Remote Sensing*, vol. 12, no. 6, Art. no. 6, Jan. 2020, doi: 10.3390/rs12060989.
- [18] P. Mittal, R. Singh, and A. Sharma, "Deep learning-based object detection in low-altitude UAV datasets: A survey," *Image and Vision Computing*, vol. 104, p. 104046, Dec. 2020, doi: 10.1016/j.imavis.2020.104046.
- [19] X. Li, J. Deng, and Y. Fang, "Few-Shot Object Detection on Remote Sensing Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2022, doi: 10.1109/TGRS.2021.3051383.
- [20] X. Sun, P. Wang, C. Wang, Y. Liu, and K. Fu, "PBNet: Part-based convolutional neural network for complex composite object detection in remote sensing imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 173, pp. 50–65, Mar. 2021, doi: 10.1016/j.isprsjprs.2020.12.015.
- [21] X. Zhang et al., "Geospatial Object Detection on High Resolution Remote Sensing Imagery Based on Double Multi-Scale Feature Pyramid Network," *Remote Sensing*, vol. 11, no. 7, Art. no. 7, Jan. 2019, doi: 10.3390/rs11070755.
- [22] C. Chen, W. Gong, Y. Chen, and W. Li, "Object Detection in Remote Sensing Images Based on a Scene-Contextual Feature Pyramid Network," *Remote Sensing*, vol. 11, no. 3, Art. no. 3, Jan. 2019, doi: 10.3390/rs11030339.
- [23] J. Yan, H. Wang, M. Yan, W. Diao, X. Sun, and H. Li, "IoU-Adaptive Deformable R-CNN: Make Full Use of IoU for Multi-Class Object Detection in Remote Sensing Imagery," *Remote Sensing*, vol. 11, no. 3, Art. no. 3, Jan. 2019, doi: 10.3390/rs11030286.
- [24] Q. Ming, L. Miao, Z. Zhou, and Y. Dong, "CFC-Net: A Critical Feature Capturing Network for Arbitrary-Oriented Object Detection in Remote-Sensing Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2022, doi: 10.1109/TGRS.2021.3095186.
- [25] X. Lu, J. Ji, Z. Xing, and Q. Miao, "Attention and Feature Fusion SSD for Remote Sensing Object Detection," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–9, 2021, doi: 10.1109/TIM.2021.3052575.
- [26] K. Fu, Z. Chang, Y. Zhang, G. Xu, K. Zhang, and X. Sun, "Rotation-aware and multi-scale convolutional neural network for object detection in remote sensing images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 161, pp. 294–308, Mar. 2020, doi: 10.1016/j.isprsjprs.2020.01.025.
- [27] J. Pang, C. Li, J. Shi, Z. Xu, and H. Feng, "Fast Tiny Object Detection in Large-Scale Remote Sensing Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 8, pp. 5512–5524, Aug. 2019, doi: 10.1109/TGRS.2019.2899955.
- [28] C. Li, C. Xu, Z. Cui, D. Wang, T. Zhang, and J. Yang, "Feature-Attentioned Object Detection in Remote Sensing Imagery," in *2019 IEEE International Conference on Image Processing (ICIP)*, Taipei, Taiwan: IEEE, Sep. 2019, pp. 3886–3890. doi: 10.1109/ICIP.2019.8803521.
- [29] X. Sun, P. Wang, C. Wang, Y. Liu, and K. Fu, "PBNet: Part-based convolutional neural network for complex composite object detection in remote sensing imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 173, pp. 50–65, Mar. 2021, doi: 10.1016/j.isprsjprs.2020.12.015.
- [30] X. Jie, S. U. O. S. A. Technology, Y. Zheng, C. Dong-Ye, P. Wang, and M. Yasir, "Improved YOLOv5 Network Method for Remote Sensing Image Based Ground Objects Recognition," In Review, preprint, Feb. 2022. doi: 10.21203/rs.3.rs-1224458/v1.
- [31] Y. Zhang, L. Guo, Z. Wang, Y. Yu, X. Liu, and F. Xu, "Intelligent Ship Detection in Remote Sensing Images Based on Multi-Layer Convolutional Feature Fusion," *Remote Sensing*, vol. 12, no. 20, p. 3316, Oct. 2020, doi: 10.3390/rs12203316.

# AIRA-ML: Auto Insurance Risk Assessment- Machine Learning Model using Resampling Methods

Ahmed Shawky Elbhrawy<sup>1</sup>, Mohamed A. Belal<sup>2</sup>, Mohamed Sameh Hassanein<sup>3</sup>

Business Information System Department-Faculty of Commerce and Business Administration, Helwan University  
Cairo, Egypt<sup>1</sup>

Professor, Computer Science Department-Faculty of Computers and Artificial Intelligence, Helwan University  
Cairo, Egypt<sup>2</sup>

Integrated Thebes Institutes for Computing & Management Science, Cairo, Egypt<sup>3</sup>

**Abstract**—Predicting underwriting risk has become a major challenge due to the imbalanced datasets in the field. A real-world imbalanced dataset is used in this work with 12 variables in 30144 cases, where most of the cases were classified as "accepting the insurance request", while a small percentage classified as "refusing insurance". This work developed 55 machine learning (ML) models to predict whether or not to renew policies. The models were developed using the original dataset and four data-level approaches resampling techniques: random oversampling, SMOTE, random undersampling, and hybrid methods with 11 ML algorithms to address the issue of imbalanced data (11 ML× (4 resampling techniques + unbalanced datasets) = 55 ML models). Seven classifier efficiency measures were used to evaluate these 55 models that were developed using 11 ML algorithms: logistic regression (LR), random forest (RF), artificial neural network (ANN), multilayer perceptron (MLP), support vector machine (SVM), naive Bayes (NB), decision tree (DT), XGBoost, k-nearest neighbors (KNN), stochastic gradient boosting (SGB), and AdaBoost. The seven classifier efficiency measures namely are accuracy, sensitivity, specificity, AUC, precision, F1-measure, and kappa. CRISP-DM methodology is utilised to ensure that studies are conducted in a rigorous and systematic manner. Additionally, RapidMiner software was used to apply the algorithms and analyze the data, which highlighted the potential of ML to improve the accuracy of risk assessment in insurance underwriting. The results showed that all ML classifiers became more effective when using resampling strategies; where Hybrid resampling methods improved the performance of machine learning models on imbalanced data with an accuracy of 0.9967 and kappa statistics of 0.992 for the RF classifier.

**Keywords**—Risk assessment; machine learning; imbalanced data; rapid miner; CRISP-DM methodology

## I. INTRODUCTION

Insurance underwriting is a critical process that assesses and selects risks. In exchange for a premium payment, an insurance company agrees to compensate the insured for financial losses under the terms of a contract between the person or organization and the insurer. Insurance is a vital risk management tool that protects individuals and businesses from unforeseen occurrences that could cause financial losses. Car insurance is one of the most important types of insurance, as it protects owners and drivers financially from a variety of dangers and uncertainties [1].

Risk assessment in the insurance sector is a critical process for evaluating the likelihood and severity of potential losses or damages for a specific policyholder. In auto insurance, risk assessment considers several factors that may increase the likelihood of an accident, such as the driver's age, driving record, vehicle type, and location. Risk assessment is essential in the world of auto insurance for accurately pricing policies and ensuring financial viability [2].

The global usage-based auto insurance market is projected to grow from \$57.86 billion in 2023 to \$174.33 billion by 2030, at a compound annual growth rate (CAGR) of 17.1%. This growth is being driven by the increasing adoption of usage-based insurance (UBI) programs by consumers, as well as the growing availability of telematics devices that can collect the data needed to calculate UBI premiums. The total direct written premium for private passenger auto insurance in the United States was \$247.1 billion in 2020, which underscores the significant size of the car insurance industry and the importance of risk assessment in ensuring that insurance companies can cover losses and remain financially stable [3] [4].

Machine learning (ML) has been used to improve the accuracy and effectiveness of risk assessment in auto insurance. ML algorithms are trained on large amounts of data to identify patterns and trends that would be difficult to find using traditional methods. This information can then be used to make more accurate predictions about the likelihood of an accident occurring, which can help insurers to price their policies more accurately and make better underwriting decisions [5].

Imbalance learning is a long-standing challenge in machine learning. In the context of auto insurance risk assessment, the majority class would be the group that reflects the majority of risks. The minority class, on the other hand, would be the group that makes up less of the total, such as policyholders who are denied insurance renewal. This category may have very little data. As a result, the data distribution across dataset classifications is often inconsistent in real-world settings. To improve the reliability of risk assessment, it is necessary to correct erroneous data. Data imbalances can be addressed using resampling techniques [6].

RapidMiner Studio is a data science platform that provides a graphical user interface for designing and deploying ML

models. It enables users to preprocess data, build ML models, and apply reshaping techniques for unbalanced data. RapidMiner supports a wide range of ML algorithms and provides tools for evaluating model performance and selecting the best model. It is an important tool for handling ML algorithms and applying reshaping techniques for unbalanced data because it provides a user-friendly interface for implementing these techniques without the need for advanced programming skills [7] [8].

The motivation for this paper is the need for accurate risk assessment in auto insurance. Risk assessment is critical for accurately pricing policies and ensuring the financial viability of insurance companies. However, traditional risk assessment methods are often inaccurate due to imbalanced datasets, which makes it difficult to predict the risk of underwriting a new policy. Machine learning (ML) techniques have been proposed to improve accuracy, but they are also susceptible to data imbalance problems.

This paper proposes a new approach to address the challenge of imbalanced data in auto insurance datasets by using resampling techniques to create a more balanced dataset. This could lead to more accurate risk predictions and better pricing decisions for insurance companies. Additionally, the proposed approach could help reduce the risk of losses and automate the risk assessment process, thereby improving underwriting efficiency.

The paper is divided into five sections. Section I introduces the topic, followed by Section II, which reviews the related work. Section III discusses the methodology, which separately describes the phases of the proposed model. Section IV, Results and Discussion presents the final steps of the methodology and its results. Section V presents the conclusion, and discusses the future research directions.

## II. RELATED WORK

In this section, the current work will review previous research efforts in risk assessment, claim prediction, and the use of machine learning algorithms and resampling methods. In the study of [9], the authors investigated the use of data mining tools and methods to develop models for analyzing risk levels for the Ethiopian Insurance Corporation (EIC). The study found that a decision tree model achieved an accuracy rate of 0.75 in classifying 3100 policies, while a neural network model achieved an accuracy rate of 58. And in [10], they investigated the use of telematics data to predict accident claims. They compared the effectiveness of XGBoost and LR methods. LR was found to be a suitable model for this task because it is interpretable and has good predictive performance with accuracy rate of 0.8397. XGBoost requires more effort to interpret and requires several model-tuning strategies to match the predictive performance of logistic regression. And in [11], the authors aimed to classify participants in the insurance renewal process to help companies reduce the claim ratio by being more selective in approving them. The proposed method involved classifying insurance participants' data using 3,803 datasets with four attributes and five algorithms to find significant features when generating the model. The study found that the decision tree (DT) algorithm was the most accurate, with an accuracy rate of 0.9540. The DT algorithm

also showed that the most significant feature in defining prospective company assessment was the average age. And in [12], the authors used bootstrapping for resampling to evaluate two classifiers, RF and SVM, using four metrics: Accuracy, Precision, Recall, and F-measure. The experimental results showed that the two classifiers scored an overall accuracy of 0.9836 and 0.9817, respectively.

And in [13], the authors investigated the use of machine learning techniques by auto insurance companies to analyse large amounts of insurance-related data and forecast claim incidence. They applied a variety of machine learning techniques, including LR, XGBoost, RF, DT, NB, and K-NN. They also applied the random over-sampling technique to address the problem of unbalanced data. The results of the study showed that the RF model outperformed all other approaches with an accuracy of 0.8677. And in [14], they developed 32 machine learning models using various data-level approaches to address this challenge. The study found that the AdaBoost classifier with oversampling and the hybrid method had the most accurate predictions. The study concluded that the AdaBoost classifier, using oversampling or the hybrid process, can generate more accurate models for analyzing imbalanced data in the insurance industry than other models.

## III. METHODOLOGY AND PROPOSED MODEL

CRISP-DM (Cross Industry Standard Process for Data Mining) methodology is adopted to develop the proposed model called AIRA-ML (Auto Insurance Risk Assessment-machine learning) shown in Fig. 1 for predicting risk assessment in car insurance. This methodology is used to ensure that the model is developed correctly and that it meets the needs of the business [15]. AIRA-ML consists of six phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment [15] that will be explain in this sections and the following two sections in details:

### A. Business Understanding

This phase focuses on understanding business needs, which is then used to create an accurate predictive model using machine learning techniques. This phase consists of the following two sub phases:

1) *Determine business objective*: Classify customers using historical data such as insured data, vehicle data, and claims data for a deeper understanding of the data due to its different sources as shown in Table I. The Remark column in Table I provides additional insights such as relationships between independent variables were determined by preliminary examination and expert opinions to identify the dependent variable.

2) *Determine machine learning goals*: ML techniques in auto insurance can provide valuable information but face challenges like risk assessment using unbalanced data. Thus, this work aims to illustrate the effects of imbalanced data and select the most effective resampling technique.



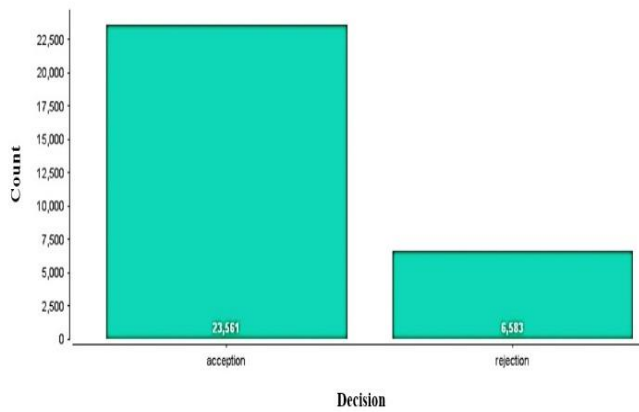


Fig. 1. Imbalanced data in the dataset.

TABLE I. SUMMARY OF THE ATTRIBUTES USED WITH THEIR DESCRIPTION

No.	Factors Group	Features Name	Description	Remark
1	Insured data	Age	Age of Policyholder	Independent
2		Location	Address	
3		Job	Class of business	
4		Hstatus	Health status	
5		Qualification	Educational Qualification	
6	Vehicle data	Make	Make of vehicle	
7		Model	Model of vehicle	
8		Body	Body type of vehicle	
9		Cc	Horsepower or CC	
10		Vage	Age of vehicle	
11		Use	Sub class of business (based on the purpose of use of vehicle)	
12		Availparts	Availability of spare parts	
13	Mileage	Mileage		
14	Claim data	Premium	Premium (averaged)	
15		Tclaimc	Total claim cost	
16		Insurance Val	Insurance Value	
17		Category	The level of probability of risk	
18	Decision	Decision	The decision to accept or refuse a policy	

**B. Data Understanding**

This phase focuses attention on finding, gathering, and analyzing the data sets that are used in AIRA-ML model phases. This phase is composed of four main tasks:

1) *Collect initial data*: a real-life data from an Egyptian car insurance company's policy and claims database were used, as well as manual formats for collecting vehicle and owner information during underwriting and claim requests.

2) *Describe the data*: The features in the dataset are 18, as shown in Table I, and 30144 records for insurance policy renewal that construct the used dataset, where insurance policy for 23561 client were renewed and 6583 were refused as shown in Fig. 2. This figure shows an imbalanced dataset problem that this work will address.

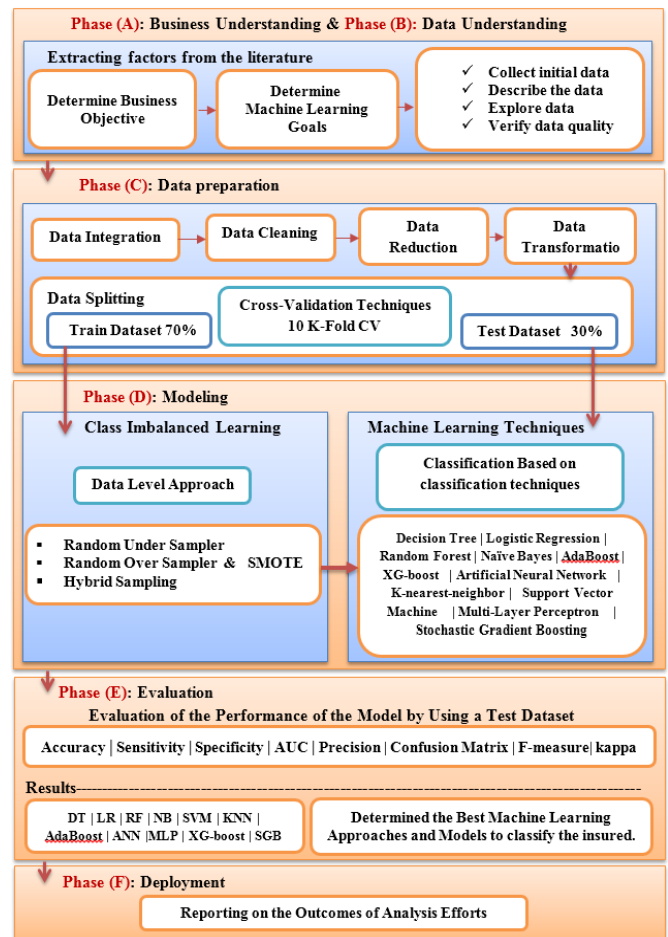


Fig. 2. AIRA-ML model.

3) *Explore data*: It is possible to infer remark column to explains that relationships between the independent variables and determine the dependent variable through preliminary examination and expert opinions. The insurance policies were classified into one of three possible categories risk (low, medium, or high) based on an annual assessment made by the insurance company. 17 features were used to classify the policies, including "accept the insurance request" and "refusal of insurance" as illustrated in Table I.

4) *Verify data quality*: This process plays a crucial role in enhancing the quality and integrity of the data, through checking out data completeness and correctness to enable more effective data-driven decision-making and insights extraction for the AIRA-ML model.

**C. Data Preparation**

Data preparation involves cleaning, transformation, feature engineering, integration, reduction, and splitting the organizing raw data for ML algorithm, which form training of classification model [8] [16]. These processes were conducted through a data science platform offering tools called RapidMiner that will be also used throughout all phases of AIRA-ML model.

1) *Data cleaning*: Data cleaning is a critical process that was performed on identifying and correcting or removing errors, inconsistencies, and inaccuracies within a dataset. By eliminating duplicate entries, handling missing values, correcting formatting issues, and dealing with outliers. In order to ensure that the dataset is accurate, reliable, and ready for further analysis.

2) *Data transformation*: The data transformation is performed on the feature that captures whether a car is used privately or commercially, which determines a specific feature. To facilitate analysis, the "Nominal to Numerical" operator is employed to convert categorical feature into numerical values. For example, in Table I, "private" is denoted as (1) and "commercial" as (0) for car usage. This conversion is applied to all categorical features in the analysis and in integral sub-process step within the AIRA-ML model, as depicted in Fig. 3. This ensures that the knowledge obtained from this transformation can be effectively utilized across the categorical features.

3) *Data integration*: This step "Join" combines data from different sources, as in the Factors Group in Table I, collecting vehicle and owner information during underwriting and claim requests. As shown in Fig. 3.

4) *Data reduction*: The "Select Attributes" selects only the most relevant features for analyzing relationships between the independent variables and determining the dependent variable through preliminary examination and expert opinions, as shown in Table I, and this can be determined through the RapidMiner as shown in Fig. 3.

5) *Data splitting*: The "Split Data" divides datasets into training and testing sets in the AIRA-ML model, splitting the data into 30% for testing and 70% for training [14], improving model performance and accuracy by ensuring data format and relevant features. This comprehensive tool helps organize raw data for machine learning training model as illustrated in Fig. 3[8].

#### D. Modeling

1) *Imbalanced data and a data-level approach*: Learning from imbalanced data is a challenging problem in machine learning, as real-life datasets often have imbalanced class distributions where a minority class has fewer samples than the majority class. As illustrated in Fig 2. This leads to biased results in standard machine learning algorithms. Minority classes may have more critical information and higher value, making it crucial to distinguish them. To overcome the bias, various techniques have been proposed in the field of imbalanced learning [16] [17].

Unbalanced data has a big impact on classification algorithm performance [8]. So this paper discusses resampling techniques like Random OverSampler, Random

UnderSampler, and SMOTE to address data imbalance and improve machine learning algorithms' performance. The authors compare these techniques and suggest improvements in classification algorithm performance.

a) *Data level methods*: The data-level approach involves oversampling and undersampling techniques to maintain balance between classes before classification [18]. And this is what it was applied for in the AIRA-ML model, as shown in Fig 1, and the implementation is illustrated in Fig. 3.

- Under-Sampling Methods

Under-sampling, also known as random under-sampling, is a technique to address unbalanced data by removing cases from the majority class of the training dataset [18].

- Over-sampling methods

Random Over-Sampling is a bootstrap-based technique for binary classification in imbalanced classes. It creates synthetic samples using conditional density estimation, working with continuous and categorical data. The technique maintains constant sample diversity without creating new samples [19].

SMOTE (Synthetic Minority Oversampling Technique) is a method of oversampling that creates fresh minority samples by mixing two minorities with one of their K nearest neighbours [6].

- Hybrid methods

Hybrid methods are a mixture of over-sampling and under-sampling methods at the data level. Hybrid sampling combines oversampling and undersampling to increase minority class numbers while decreasing majority class numbers.

#### 2) Implementing the data-level methods

a) *RUS (Random Under Sampling)*: RUS was implemented by simply choosing a random sample from the acceptance class ("majority class") that corresponds to the number of samples from the rejection class ("minority class"). Random undersampling of the majority class was accomplished as illustrated in Fig. 3.

b) *ROS (Random Over Sampling)*: To implement ROS, a procedure that produces an equal number of replicants of the minority class as samples of the majority class was developed. The bootstrapping operator in step resampling Techniques, as shown in Fig. 3. By doing several resamples of the original dataset using random oversampling of the rejection class with replacement, this operator generates a bootstrapped sample from the original dataset.

c) *The ROS/RUS (Random Over Sampling / Random Under Sampling)*: The aforementioned approach is modified in the ROS/RUS, where a sample size of the acceptance class and rejection class is selected in the step with resampling techniques as illustrated in Fig. 3.

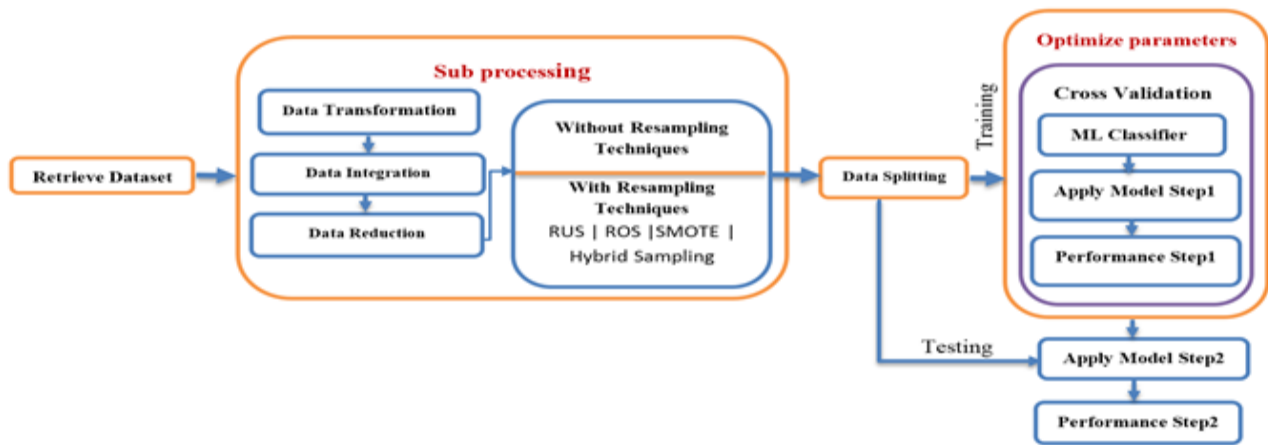


Fig. 3. The Process performed in RapidMiner with AIRA-MI Model.

d) *SMOTE (Synthetic Minority Oversampling Technique)*: The SMOTE technique was used on the dataset until the ratio of minority samples to majority samples is equal as illustrated in Fig. 3.

3) *Modeling techniques*: In the AIRA-ML model, 11 machine learning classifiers are used, as previously stated in Fig. 1. The AIRA-ML model incorporates 10-fold cross-validation to ensure fair comparisons by selecting machine learning parameters for each model. RapidMiner parameter optimization feature is employed to determine the ideal values for the chosen parameters, resulting in optimal outcomes across various machine learning models. The 10-fold cross-validation method is utilized by dividing the data into two groups: approximately 30% for test data and 70% for training data. Models are developed using the training data and evaluated using the test data.

#### IV. RESULTS AND DISCUSSION

##### A. Evaluation

Evaluation techniques calculate the effectiveness of classifiers in selecting the best applied model. Accuracy alone may not solve classification problems due to bias in majority class results, especially in imbalanced data [13] [16]. Consequently, Confusion Matrix evaluation criteria are used for measuring accuracy, sensitivity, specificity, precision, recall, AUC stands for "Area Under the Receiver Operating Characteristic Curve". It is a metric used to evaluate the performance of a binary classifier, and F-Measure, are utilized, beside Kappa Statistics to accurate more precise insurance policy renewal acceptance/ rejection.

1) *Confusion matrix*: A confusion Matrix is employed in binary classification problems to ensure accurate predictions of class outputs and facilitate the comparison between predicted and actual classes [16]. The matrix, as shown in Table II, displays the proportions of correctly classified samples (TP) and incorrectly classified samples (FP/FN). In other words, TP signifies renewal acceptance, while TN indicates rejection.

TABLE II. CONFUSION MATRIX

	<i>Predicted Positive</i>	<i>Predicted Negative</i>
<i>Actual Positive</i>	True positive (TP)	False negative(FN)
<i>Actual Negative</i>	False positive (FP)	True negative (TN)

Accuracy is the fraction of predictions rate for insurance state are correct which is calculated in "(1)", Sensitivity is the true positive ratio of positively classified cases that are actually positive for rejected insurance policies. Which is calculated using "(2)" Specificity is true negative rate of negatively classified cases that are actually negative for accepted insurance policies which is calculated using "(3)" [13] [20].

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FP} + \text{TN} + \text{FN}) \quad (1)$$

$$\text{Sensitivity} = \text{TP} / (\text{TP} + \text{FN}) \quad (2)$$

$$\text{Specificity} = \text{TN} / (\text{FP} + \text{TN}) \quad (3)$$

The metric "(4)" is the percentage of the relevant outcomes that assess the reliability of the classification and determine the appropriateness of acceptance or rejection decision. A recall "(5)" is a measurement of how many positive instances were correctly identified as positive, especially for imbalanced datasets [13] [20].

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \quad (4)$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) \quad (5)$$

The F-measure also known as the F1 score "(6)" is a measure of a model's performance that takes into account the averaging of precision and recall.

$$\text{F-measure} = (2 * \text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (6)$$

2) *Kappa statistics*: Kappa statistics play a valuable role in assessing prediction success for both class acceptance and rejection, considering factors beyond accuracy. This is particularly important when dealing with datasets that exhibit significant imbalance class. Kappa "(7)" takes into account the agreement between model predictions and actual labels, providing a measure of agreement that goes beyond accuracy alone [13].

$$K=(Pr(a)-Pr(e))/(1-Pr(e)) \quad (7)$$

Where:

Pr(a) represents the observed agreement between the raters, which is the proportion of cases where the raters agree.

Pr(e) represents the expected agreement between the raters by chance.

Results of 11 machine learning classifiers on unbalanced data without any resampling models applied are shown in Fig. 4. The result highlights the highest accuracy rate in the classifiers with DT achieving an accuracy of 0.9015, because of machine learning techniques often ignore the minority class (rejection class) and allocate most cases to the majority class (acceptance class). Moreover, a direct proportion relationship can be seen between accuracy and specificity, whenever accuracy is high, the specificity of 0.949 is high as well due to the model consistently predicting the majority class, but reduced sensitivity to 0.6333. Precision is low at 0.8768 because the model frequently predicts the minority class incorrectly. The F-measure is also low at 0.1482 due to the combination of low precision and recall of 0.6333. Among the classifiers, kappa is low at 0.3267 because the model is not very good at predicting the minority class, and AUC is misleading because it is high and the model is not very good at predicting the minority class. On the other hand, the MLP classifier had the lowest performance measures (accuracy = 0.6888, sensitivity = 0.0825, specificity = 0.92121, AUC = 0.5021, precision = 0.7294, F1-measure = 0.14823, and kappa = 0.0871).

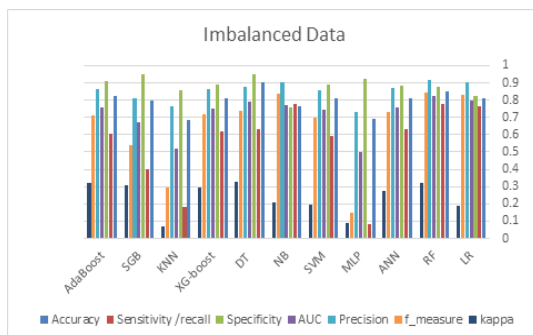


Fig. 4. Classifiers metrics for imbalanced data.

Fig. 5 shows the results of machine learning algorithms on imbalanced data that has been resampled using random oversampling. The accuracy results are not significantly improved, with the SVM classifier achieving an accuracy of 0.8736. This is understandable given that most models predict with poorer accuracy on balanced data, as they consider all classes at the same time. Accuracy is a simple metric to understand, but it overlooks several important factors that must be considered when evaluating a classifier's output. Therefore, we used additional metrics. Additionally, it is noteworthy that the sensitivity for all models with imbalanced data is lower than the sensitivity for balanced data created by random oversampling. This is because random oversampling does not address the class imbalance problem in a principled way. It simply increases the number of minority class samples lead to overfitting. The specificity for the SVM classifier is 0.8601,

with a high F-measure of 0.9366. This is due to the combination of high precision 0.9662 and recall 0.9087. The kappa is low 0.4514 because the model is not very good at predicting the minority class. The AUC is also not very good 0.8841. The KNN classifier has the lowest performance measures. The accuracy is 0.5563, the sensitivity is 0.4594, the specificity is 0.5934, the AUC is 0.5261, the precision is 0.7444, the F-measure is 0.5681, and the kappa is 0.112.

Overall, the results show that random oversampling is not an effective way to address the class imbalance problem in machine learning. Other principled approaches, such as SMOTE or hybrid method, are applied to improve the performance of imbalanced data.

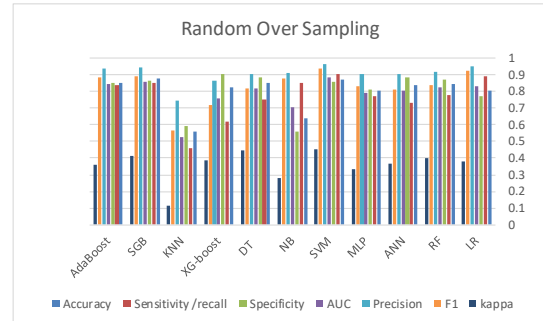


Fig. 5. Classifiers metrics using the ROS method.

The results of machine learning algorithms on imbalanced data that has been resampled using random undersampling are shown in Fig. 6. The high precision for many classifiers, such as random forest (RF) 0.9823 and Artificial Neural Network ANN 0.9505, is because the model frequently predicts the minority class correctly. The F-measure is also high for RF 0.9669 due to the combination of high precision and recall 0.9521, the kappa is high with comparison with Fig. 4 and Fig. 5. Moreover, such as ANN 0.4514, the model is very good at predicting the minority class. The AUC is 0.8391, and accuracy is 0.8713, specificity is 0.799. On the other hand, the KNN classifier has the lowest performance measures. With accuracy of 0.5001, sensitivity of 0.6913, specificity of 0.4268, AUC of 0.5591, precision is 0.7833, F1-measure of 0.7344, and kappa of 0.254.

Additionally, Random undersampling is a technique that reduces the number of majority class samples in a dataset. This can help to improve the performance of machine learning models on imbalanced data, as it reduces the bias towards the majority class. However, the results show that random undersampling can be an effective way to address the class imbalance problem in machine learning.

The best performance with SMOTE is DT classifier as shown in Fig. 7 With an accuracy of 0.8575, sensitivity of 0.9521, specificity of 0.8212, AUC of 0.9521, precision of 0.8871, F1-measure of 0.9184, and a kappa of 0.483, while the lowest performance at SMOTE is the KNN classifier with an accuracy of 0.5844, sensitivity of 0.4159, specificity of 0.6490, AUC of 0.4159, precision of 0.5321, F1-measure of 0.4668, and kappa of 0.225. The result of the SMOTE resampling method has shown the improvement for sensitivity and specificity of the model. For example, the ANN classifier of

the SMOTE results has significantly improved with a sensitivity of 0.9087 and a specificity of 0.8212; in comparison with the ANN classifier without SMOTE, which has a sensitivity of 0.6333, and a specificity of 0.8823 as shown in Fig. 4. The SMOTE resampling method is also improving the performance of other classifiers, such as the DT classifier and the KNN classifier. Overall, the results show that the SMOTE resampling method is an effective way to improve the performance of machine learning models on imbalanced data, which leads to a more accurate and reliable model.

The hybrid resampling methods are more effective than random oversampling or undersampling alone. This is because they are able to create a more balanced dataset without overfitting the models.

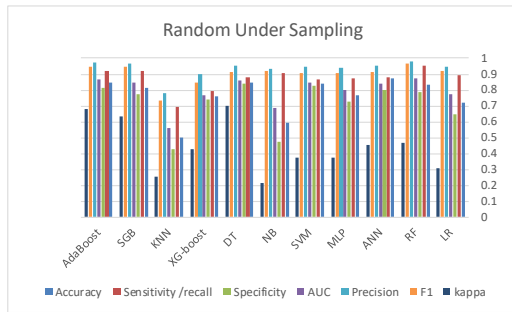


Fig. 6. Classifiers metrics using the RUS method.

The results are shown in Fig. 8. The random forest (RF) classifier with hybrid resampling has the best performance, with an accuracy of 0.9967, an AUC of 0.994, a precision of 0.9977, an F-measure of 0.9975, and a kappa of 0.992. This means that the RF classifier is very accurate in predicting whether accept or reject the renewal of the policies. It also has the smallest gap between sensitivity of 0.9977 and specificity of 0.9959, which is an important performance indicator while

the MLP classifier has the lowest performance, with an accuracy of 0.5242, a sensitivity of 0.5029, a specificity of 0.5323, an AUC of 0.5171, a precision of 0.7388, an F-measure of 0.5984, and a kappa of 0.0497.

Table III compare the purposed model (AIRA-ML model) with earlier studies that used other model and resampling technique. The data was preprocessed in both its balanced and unbalanced states to improve the accuracy of training result and the effectiveness of machine learning algorithms.

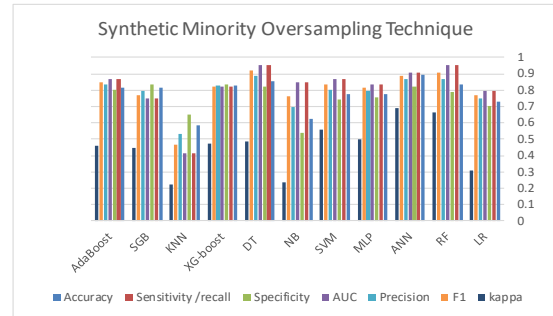


Fig. 7. Classifiers metrics using the SMOTE method.

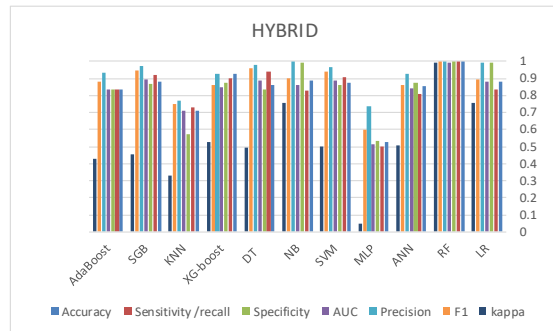


Fig. 8. Classifiers metrics using the hybrid method.

TABLE III. COMPARATIVE RESULTS FOR NEW APPROACH PERFORMANCE AGAINST RELATED WORKS

Related work	Resampling Methods	Machine Learning Techniques	The best model	Performance Measures						
				Accuracy	Sensitivity / Recall	Specificity	AUC	Precision	F-measure	Kappa
[9]	Random Sampling, SMOTE	KNN, DT	DT	0.75	0.749	×	×	×	×	×
[10]	×	LR, XGBoost	XGBoost	0.8397	0.0790	0.9022	×	×	×	×
[11]	×	NB, SVM, LR, NN, DT	DT	0.9540	×	×	×	×	×	×
[12]	Bootstrapping	SVM, RF	RF	0.9836	0.9471	×	×	0.9515	0.9490	×
[13]	ROS	LR, C5.0, J48, XGBoost, DT, NB, K-NN, RF	RF	0.8677	0.9717	0.71	0.840	0.9429	0.8101	0.7117
[14]	ROS, RUS, Hybrid, SMOTE	C5.0, C4.5, CART, Bagged CART, RF, XGBoost, SGB, AdaBoost	AdaBoost	0.9940	0.9294	0.9982	×	×	×	×
AIRA-ML Model	ROS, RUS, SMOTE, Hybrid	LR, ANN, MLP, SVM, NB, DT, XG-boost, KNN, SGB, AdaBoost, RF	RF	0.9967	0.9977	0.9959	0.8701	0.9977	0.9975	0.992

×- Not used;

Finally, the hybrid approach performs is much better than the results of the earlier research. This approach, which is very accurate with predicting the decision of policies renewal at the same time as an essential performance indicator, that has the smallest gap between sensitivity and specificity. AUC, F1-score, and Kappa statistics are used as other Performance Measures to ensure that the model is effective as a trustworthy instrument for risk assessment and insurance policies activities in the insurance industry.

### B. Deployment

The deployment of the model is beyond the scope of this work and is the responsibility of the insurance company. The purpose of the model is to increase knowledge of the data, and the knowledge gained will need to be organized and presented in a way that the customer can use it. The research in this paper was conducted primarily for academic purposes, but the results can be used by the financial sector to apply machine learning technology to improve their business practices and gain a competitive edge.

However, the research has identified a task that needs more consideration in future work. Proper handling and concern for information are strongly recommended in data mining research.

## V. CONCLUSION AND FUTURE WORK

The insurance industry faces a significant challenge in predicting risk assessment in insurance policies. Thus, this paper proposes an accurate predictive model using machine learning (ML) and resampling techniques to assist insurance companies in making better pricing decisions. The results demonstrate that ML can be used to create an accurate predictive model for auto insurance risk, which can improve insurance acceptance and pricing decisions. Additionally, the results demonstrate that ML can be effective in addressing data imbalance problems in the auto insurance sector. The hybrid resampling technique outperformed all other resampling techniques, achieving an accuracy of 99.6% for the random forest (RF) classifier. This suggests that the hybrid resampling method is a promising approach for dealing with class imbalance problems in ML.

Further research is required to compare the efficiency measures using various datasets from various fields to prove the prediction efficiency of a random forest classifier with resampling methods to solve the imbalanced data problem. And future work may be done in the following directions: Using hybrid resampling techniques to improve comparison and performance with machine learning classifiers, apply this methodology to other sectors of insurance or any other sector with the same problem, which has an imbalance of data.

### REFERENCES

- [1] Et. al., Nadia Yas. 2021. "Implications of Compulsory Car Accident Insurance Comparative Study." Turkish Journal of Computer and Mathematics Education (TURCOMAT) 12 (2): 2410–20. <https://doi.org/10.17762/turcomat.v12i2.2052>.
- [2] Radic, M., P. Herrmann, P. Haberland, and Carla R. Riese. 2022. "Development of a Business Model Resilience Framework for Managers and Strategic Decision-Makers." Schmalenbach Journal of Business Research 74 (4). <https://doi.org/10.1007/s41471-022-00135-x>.
- [3] F. B. Insights, "Automotive Usage Based Insurance Market Size | Growth [2028]," Jun 2023. [Online]. Available: <https://www.fortunebusinessinsights.com/automotive-usage-based-insurance-market-104103>. [Accessed 1 8 2023].
- [4] Drakulevski, Ljubomir, and Tamara Kaftandzieva. 2021. "Risk Assessment Providing Solid Grounds For Strategic Management In The Insurance Industry." European Scientific Journal ESJ 17 (15): 38–56. <https://doi.org/10.19044/esj.2021.v17n15p38>.
- [5] Rawat, Seema, Aakankshu Rawat, Deepak Kumar, and A. Sai Sabitha. 2021. "Application of Machine Learning and Data Visualization Techniques for Decision Support in the Insurance Sector." International Journal of Information Management Data Insights 1 (2): 1–15. <https://doi.org/10.1016/j.ijimei.2021.100012>.
- [6] T Wongvorachan, Tarid, Surina He, and Okan Bulut. 2023. "A Comparison of Undersampling, Oversampling, and SMOTE Methods for Dealing with Imbalanced Classification in Educational Data Mining." Information (Switzerland) 14 (1): 1–15. <https://doi.org/10.3390/info14010054>.
- [7] Andry, Johanes Fernandes, Henny Hartono, Honni, Aziza Chakir, and Rafael. 2022. "Data Set Analysis Using Rapid Miner to Predict Cost Insurance Forecast with Data Mining Methods." Journal of Human University Natural Sciences 49 (6): 167–75. <https://doi.org/10.55463/issn.1674-2974.49.6.17>.
- [8] Madyatmadja, Evaristus Didik, Samuel Imanuel Jordan, and Johanes Fernandes Andry. 2021. "Big Data Analysis Using Rapidminer Studio to Predict Suicide Rate in Several Countries." ICIC Express Letters, Part B: Applications 12 (8): 757–64. [10.24507/icicelb.12.08.757](https://doi.org/10.24507/icicelb.12.08.757).
- [9] Wuyu, Sisay, and Patrick Cerna. 2018. "Risk Assessment Predictive Modelling in Ethiopian Insurance Industry Using Data Mining." Software Engineering 6 (4): 121–27. <https://doi.org/10.11648/j.se.20180604.13>.
- [10] Pesantéz-Narvaez, Jessica, Montserrat Guillen, and Manuela Alcañiz. 2019. "Predicting Motor Insurance Claims Using Telematics Data—XGboost versus Logistic Regression." Risks 7 (2). <https://doi.org/10.3390/risks7020070>.
- [11] Utomo, Dedy, Noperida Damanik, and Indra Budi. 2021. "Classification on Participants Renewal Process in Insurance Company: Case Study PT XYZ." In 2021 9th International Conference on Information and Communication Technology (ICoICT), 576–81. IEEE. <https://doi.org/10.1109/ICoICT52021.2021.9527479>.
- [12] Alamir, Endalew, Teklu Urgessa, Ashebir Hunegnaw, and Tiruveedula Gopikrishna. 2021. "Motor Insurance Claim Status Prediction Using Machine Learning Techniques." International Journal of Advanced Computer Science and Applications 12 (3): 457–63. <https://doi.org/10.14569/IJACSA.2021.0120354>.
- [13] Hanafy, Mohamed, and Ruixing Ming. 2021. "Machine Learning Approaches for Auto Insurance Big Data." Risks 9 (2): 1–23. <https://doi.org/10.3390/risks9020042>.
- [14] Hanafy, Mohamed, and Ruixing Ming. 2021. "Improving Imbalanced Data Classification in Auto Insurance by the Data Level Approaches." International Journal of Advanced Computer Science and Applications 12 (6): 493–99. <https://doi.org/10.14569/IJACSA.2021.0120656>.
- [15] Martinez-Plumed, Fernando, Lidia Contreras-Ochando, Cesar Ferri, Jose Hernandez-Orallo, Meelis Kull, Nicolas Lachiche, Maria Jose Ramirez-Quintana, and Peter Flach. 2021. "CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories." IEEE Transactions on Knowledge and Data Engineering 33 (8): 3048–61. <https://doi.org/10.1109/TKDE.2019.2962680>.
- [16] Baran, Sebastian, and Przemysław Rola. 2022. "Prediction of Motor Insurance Claims Occurrence as an Imbalanced Machine Learning Problem." ArXiv abs/2204.0: 1–12. <https://doi.org/10.48550/arXiv.2204.06109>.
- [17] Mohammed, Roweida, Jumanah Rawashdeh, and Malak Abdullah. 2020. "Machine Learning with Oversampling and Undersampling Techniques: Overview Study and Experimental Results." 2020 11th International Conference on Information and Communication Systems, ICICS 2020, no. May: 243–48. <https://doi.org/10.1109/ICICS49469.2020.239556>.
- [18] Dasari, Siva Krishna, Abbas Cheddad, Jonatan Palmquist, and Lars Lundberg. 2022. "Clustering-Based Adaptive Data Augmentation for Class-Imbalance in Machine Learning (CADA): Additive Manufacturing

- Use Case.” *Neural Computing and Applications* 6.  
<https://doi.org/10.1007/s00521-022-07347-6>.
- [19] Le, Tuong, Minh Thanh Vo, Bay Vo, Mi Young Lee, and Sung Wook Baik. 2019. “A Hybrid Approach Using Oversampling Technique and Cost-Sensitive Learning for Bankruptcy Prediction.” *Complexity* 2019. <https://doi.org/10.1155/2019/8460934>.
- [20] Uddin, Moin, Mohd Faizan Ansari, Mohd Adil, Ripon K. Chakraborty, and Michael J. Ryan. 2023. “Modeling Vehicle Insurance Adoption by Automobile Owners: A Hybrid Random Forest Classifier Approach.” *Processes* 11 (2): 1–16. <https://doi.org/10.3390/pr11020629>.

# An MILP-based Lexicographic Approach for Robust Selective Full Truckload Vehicle Routing Problem

Karim EL Bouyahyioui, Anouar Annouch, Adil Bellabdaoui

ITM-Information Technology and Management, ENSIAS-Mohammed V University in Rabat, Morocco

**Abstract**—Full truckload (FTL) shipment is one of the largest trucking modes. It is an essential part of the transportation industry, where the carriers are required to move FTL transportation demands (orders) at a minimal cost between pairs of locations using a certain number of trucks available at the depots. The drivers who pick up and deliver these orders must return to their home depots within a given time. In practice, satisfying those orders within a given time frame (e.g., one day) could be impossible while adhering to all operational constraints. As a result, the investigated problem is distinguished by the selective aspect, in which only a subset of transportation demands is serviced. Furthermore, travel times between nodes can be uncertain and vary depending on various possible scenarios. The robustness subsequently consists of identifying a feasible solution in all scenarios. Therefore, this study introduces an MILP-based lexicographic approach to solve a robust selective full truckload vehicle routing problem (RSFTVRP). We demonstrated the proposed method's efficiency through experimental results on newly generated instances for the considered problem.

**Keywords**—Vehicle routing problem; full truckload; robust optimization; MILP-based lexicographic approach; uncertain travel time

## I. INTRODUCTION

The vehicle routing problem (VRP) is one of the various investigated combinatorial optimization problems [1]. Its tendency is due to its concrete application in the logistics and transportation domains. These fields play an essential role in the modern market economy by assuring the movement of goods from factories to customers. The FTVRP is a variation of the VRP that has previously garnered limited scientific interest. This problem has a significant application to the truckload industry, where the carriers must service FTL transportation demands (known in the literature as orders or commodities) at a minimal cost between pairs of locations utilizing an available fleet of trucks. They are given a network of sites with FTL orders to be shipped between some pairs of locations. These orders must be shipped using a certain number of trucks available at the depots. The drivers who pick up and deliver these orders must return to their home depots (domiciles) within a given time. The problem is determining the least-cost truck routes so that every order is picked up at its source and shipped to its destination. A typical truck would leave the depot, pick up a commodity, deliver the commodity, travel empty, pick up another commodity, deliver the commodity, and so on. Finally, after picking up and delivering some orders, the truck returns to its domicile. Moreover, each order must be picked between certain hours only. In other

words, every pickup must only be made between a specific pickup time window. Once the order is picked up, it must then be delivered to its destination. Depending on the specific problem, the delivery can be made at any time or within a specified delivery time window only. If a driver reaches a pickup location early, he usually has to wait until the open time. In some instances, the driver may be paid based on a certain hourly rate for staying. The integration of the time constraints in the FTL transportation problem gives rise to the FTVRP with time windows (FTVRPTW). Furthermore, trucking companies can service their clients through numerous depots, with each client assigned preferentially to one depot (multi-depot FTVRP, FTMDVRP).

On the one hand, meeting all of the aforementioned attributes may prevent the trucks from honoring all orders. Therefore, the selective feature of the problem is introduced by relaxing the requirement of servicing all transportation demands within a limited time (selective FTVRP, SFTVRP). The goal may be the maximization of the collected total profit, in which a profit is associated with each order, the minimization of the whole travelling costs, in which the assignment covers as many feasible orders as possible, or the optimization of a combination of both.

On the other hand, travel between a pair of locations might be made via multiple paths. However, the optimal one with the shortest travel time will usually be traversed. In practice, the travel time is uncertain and dependent on specific circumstances (peak traffic hours, weather conditions, accidents, and so on). As a result, this paper investigates a robust selective full truckload multi-depot vehicle routing problem with time windows (RSFTMDVRPTW) under uncertainty in transportation time as a set of discrete scenarios. Each one illustrates an eventual situation that could occur throughout the shipping period. In each scenario, we consider that a fixed number of edges connecting locations are perturbed, and the travel times along those edges differ from the ideal ones. In this study, we formulate the RSFTMDVRPTW as a mixed-integer linear programming (MILP) model under uncertainty and the multi-objective facet, which addresses two functions: an economic component to be maximized and a component related to the worst observation of the total travel time over all scenarios to be minimized. These two objectives are conflicting in that increasing profit necessitates servicing more commodities, resulting in a longer transit time.

Many papers in the literature have been devoted to introducing, formulating, and solving FTVRP variants. Among them, the SFTVRP was solved using mathematical models,



exact solvers, meta-heuristics, and hybrid methods. The novelty of the problem under consideration is that the SFTVRP is treated in a robust backhaul trucking context while considering two objectives. As a result, combining the robust and selective aspects helps to bring the model closer to reality.

The remainder of this study is organized as follows: Section II describes some of the related works. Section III then formulates the RSFTMDVRPTW, while Section IV describes our MILP-based lexicographic approach. Section V presents the experimental findings. Finally, Section VI concludes the paper and suggests some future research directions.

## II. RELATED WORKS

Over the last two decades, many researchers have studied various FTVRP variants by adding constraints to the fundamental problem to better match real-world applications. Different approaches are used to solve these variants. Interested readers are referred to [2] for a detailed review of FTVRPs. To position our study in relation to the literature, we classified the various contributions of selective FTVRP (SFTVRP) variants based on whether the routes primarily contain the FTL shipment, transportation demand selection, multiple depots, time window constraints (TWs), and uncertainty. Table I outlines the most critical SFTVRP-related works and the current study features.

Ball et al. [3] were the first to introduce the multi-depot SFTVRP (SFTMDPDP). The problem consists of constructing routes for private vehicles and subcontracting chemical product commodities to common carriers with the goal of decreasing total cost while meeting a maximum time limit on truck routes. For resolving the FTMDPDP, three heuristics are proposed: a greedy insertion strategy (GI) and two algorithms based on the route-first, cluster-second (RF-CS) technique. Wang and Regan [4] investigated an SFTVRP with time windows (SFTPDPTW) considering only loading TWs, in which the objective is to minimize the empty travelling cost while serving the maximum number of orders within their time constraints. They devised an iterative strategy for resolving the problem by employing the window-partition-based (WPB) algorithm. Miori [5] proposed a TS algorithm for solving a similar SFTVRPT without TWs and with the same later goal. Li and Lu [6] presented a hybrid GA based on improved savings for an SFTVRP with split orders and the objective of maximizing the total profit. Liu et al. [7] developed a memetic algorithm to solve an SFTVRP in collaborative transportation. Another SFTVRPT in the collaborative logistics context was presented in [8]. The authors formulated the problem as an MILP model with the objective function of minimizing the total cost. They proposed a branch-and-cut-and-price-based heuristic to solve the model. Wang et al. [9] considered an SFTVRP with heterogeneous fleet application in the petrochemical industry that involves rich features, including multiple loading locations, optional orders, and loading dock capacity

limitations. They presented an MILP mathematical model for the problem, which is solved using the commercial solver Gurobi.

Yang et al. [10], Tjokroamidjojo et al. [11], and Zolfagharinia and Houghton [12] used some rolling horizon approaches (variants of re-optimization or heuristics) to deploy dynamic SFTVRP variants. A fraction of orders to be carried in a given day become known only a short time before service is needed, truck movements are added to the system as the day advances, and orders must periodically be reassigned. Li et al. [13] addressed a dynamic SFTVRP in a collaborative context in which the carrier can dynamically select its collaborative requests based on the surplus of its transport capacity in the collaborative process. The authors proposed a mixed integer programming (MIP) model for the problem, aiming to maximize the carrier's total profits after outsourcing requests. The model is solved through CPLEX software. Annouch and Bellabdaoui [14] proposed an adaptive GA to solve the open FTMDVRPTW with split delivery (FTOVRPTWSD) in the liquefied petroleum gas (LPG) distribution industry. The FTVRP variant addressed in this study is not selective, and the robustness aspect is not considered.

In our previous studies, we investigated a mono-objective variant of the selective FTMDVRPTW (SFTMDVRPTW) in an empty return context. The objective is to maximize the total profit of selective routes. The resolution of this problem is based on the development of mathematical models, exact solvers, meta-heuristics, and hybrid methods. We described a mathematical formulation of the SFTMDVRPTW as an MILP model [15]. Numerical results on small and medium-size instances are presented using the CPLEX solver. To solve larger instances, we developed, adapted, and applied some heuristic methods: an ant colony system (ACS) [16], a genetic algorithm (GA) [17-18], and a reactive tabu search (RTS) [19].

To the best of our knowledge, while uncertainty is present and relevant, it is rarely addressed for SFTVRP variants. Hammami et al. [20] investigated an SFTVRP with uncertain clearing prices. They developed an exact non-enumerative algorithm to obtain optimal solutions for small instances and a two-phase hybrid heuristic to solve larger instances.

The main contributions of this paper can be summarized as follows:

- Formulate the RSFTMDVRPTW as an MILP model under uncertainty and the multi-objective facet, which addresses two conflicting functions: an economic component to be maximized and a component related to the worst observation of total operation time to be minimized.
- Introduce a MILP-based lexicographic method for solving the RSFTMDVRPTW.

TABLE I. POSITION OF OUR STUDY IN RELATION TO THE FTVRP LITERATURE

Reference		Constraints									Objective function	Area	Solution approach	
Authors	Year	TW	Het	MD	Cap	SD	S	D	U	type			Method	
Ball et al. [3]	1983			x			x				Min. cost	Academic	H	• RF-CS • Greedy insertion
Wang and Regan [4]	2002	x		x			x				Min. EMV	Service transportation	H	WPB
											Max. #Ord			
Yang et al. [10]	2004	x					x	x			Min. cost	Service transportation	E H	• Re-optimization • RHA
Tjokroamidjojo et al. [11]	2006	x					x	x			Min. cost	Service transportation	E H	• Re-optimization • RHA
Liu et al. [7]	2010						x				Min. cost	Academic	PBM	Memetic Algorithm
Miori [5]	2011	x					x				Min. cost	Academic	SSBM	Goal programming with TS
											Max. #Ord			
											Min. #Veh			
Li and Lu [6]	2014				x	x	x				Max. profit	Service transportation	PBM	Hybrid Genetic Algorithm
Zolfagharinia and Haughton [12]	2014	x		x			x	x			Max. profit	Service transportation	H	RHA
Li et al. [13]	2015			x			x	x			Max. profit	Service transportation	E	CPLEX
El Bouyahyiouy and Bellabdaoui [16]	2017	x	x	x			x				Max. profit	Academic	PBM	Ant Colony System
Hammami et al. [20]	2021						x		x		Max. profit	Service transportation	E H	• Branch-and-Cut • Hybrid heuristic
Wang et al. [9]	2021	x	x		x		x				Min. cost	Service transportation	E	Gurobi
El Bouyahyiouy and Bellabdaoui [17]	2022	x	x	x			x				Max. profit	Academic	E PBM	• CPLEX • Genetic Algorithm
Öner and Kuyzu [8]	2021						x				Min. cost	Academic	E	Branch-and-cut-and-price
<b>This study</b>		<b>x</b>	<b>x</b>	<b>x</b>			<b>x</b>		<b>x</b>		<b>Max. profit</b> <b>Min. worst TT</b>	<b>Academic</b>	<b>E</b>	<b>MILP-based lexicographic</b>

Note: Het: Heterogeneous Fleet; MD: Multi-Depot; Cap: Capacitated; SD: Split delivery; S: Selective; D: Dynamic; U: uncertainty; TT: Travel time; #Veh: Number of vehicles; #Ord: Number of served orders; EMV: Empty vehicle movements; E: exact; H: heuristic; SSBM: Single solution-based metaheuristic; PBM: Population based metaheuristic; RHA: Rolling horizon planning approach; RF-CS: Route-first, cluster-second.

### III. A MATHEMATICAL FORMULATION OF THE RSFTMDVRPTW

#### A. Problematic

This study addresses a variant of the full truck vehicle routing problem under uncertainty in transportation time (RSFTMDVRPTW). The position of the problem is as follows. Assume a set of  $n$  orders to be served by a fixed fleet of  $m$  trucks. Each truck  $k$  is characterized by a starting point  $D_k$ , an ending point  $A_k$ , an earliest service start date  $D_k^{min}$  and a latest service end date  $D_k^{max}$ . Each order  $O_i$  ( $i = 1, \dots, n$ ) is characterized by a collection point  $L_i$  (origin) and a delivery point  $U_i$  (destination), a profit  $p_i$  (determined on the basis of the distance between origin and destination), a loading time window  $[L_i^{min}, L_i^{max}]$ , and an unloading time window  $[U_i^{min}, U_i^{max}]$ . The travel time for each arc  $(i, j) \in \{(D_k, LU_i)\} \cup \{(LU_i, LU_j)\} \cup \{(LU_i, A_k)\} \cup \{(D_k, A_k)\}$  is described by a set of  $NS$  scenarios  $t_{ij}^\xi$  ( $\xi = 1, \dots, NS$ ), where

each scenario reflects a potential time requirement for a truck traversing arc  $(i, j)$ .

The problem consists of selecting a subset of orders to be served and assigning them to trucks, thus finding an optimal sequence of orders assigned to each truck while maximizing total profit, minimizing the worst observation of the total travel time over all scenarios and respecting availability and time window constraints. Fig. 1 depicts a solution representation for an RSFTMDVRPTW instance with two trucks and 14 orders.

#### B. A discrete Scenario-based MILP Model

In this section, we propose an MILP model of the RSFTMDVRPTW, in which a set of discrete scenarios represents the uncertain travel times. The distribution of the uncertainty parameters is assumed to be unknown. As a result, all scenarios are generated uniformly. Table II defines all data and variable notations. Next, the objectives and constraints are introduced and explained.

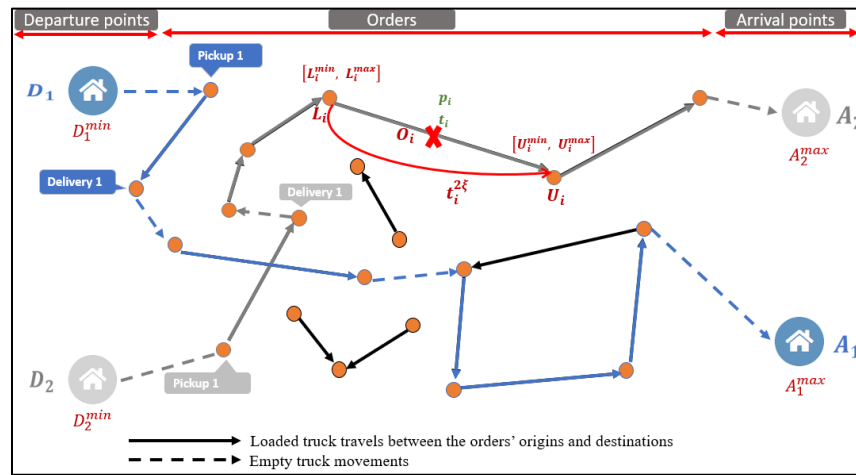


Fig. 1. An illustration of the RSFTMDVRPTW.

TABLE II. DECISION VARIABLES AND PARAMETERS OF THE MILP MODEL OF THE RSFTMDVRPTW

Notation	Meaning
$m$	Number of trucks
$D_k$	Departure depot of truck $k$
$D_k^{min}$	Earliest service start time of truck $k$
$A_k$	Arrival depot of truck $k$
$A_k^{max}$	Latest service end time of truck $k$
$n$	Number of commodities
$\{O_1, \dots, O_n\}$	Set of commodities
$L_i$	Collection point (origin) of the commodity $O_i$
$U_i$	Delivery point (destination) of the commodity $O_i$
$p_i$	profit associated with commodity $O_i$
$L_i^{min}$	Earliest time to load the commodity $O_i$
$L_i^{max}$	Latest time to load the commodity $O_i$
$U_i^{min}$	Earliest time to unload the commodity $O_i$
$U_i^{max}$	Latest time to perform the unloading of the commodity $O_i$
$NS$	Number of scenarios
$\Xi$	Set of possible scenarios
$t_i^\xi$	Travel time between collection and delivery points of the commodity $O_i$ under scenario $\xi \in \Xi$
$t_{ij}^\xi$	Empty travel time from the collection point of commodity $O_i$ to the delivery point of commodity $O_j$ under scenario $\xi \in \Xi$
$t_{oi}^{k\xi}$	Empty travel time from departure depot $D_k$ to the collection point of commodity $O_i$ under scenario $\xi \in \Xi$
$t_{i,n+1}^{k\xi}$	Empty travel cost from the delivery point of commodity $O_i$ to the arrival depot $A_k$ under scenario $\xi \in \Xi$
$M$	A big number
<b>Decision variables</b>	
$x_{ij}^k$	Binary decision variable that indicates whether the truck $k$ visits commodity $O_j$ immediately after commodity $O_j$
$t_{i,L}^{k\xi}$	Start time of the loading of commodity $O_i$ on truck $k$ under scenario $\xi \in \Xi$
$t_{i,U}^{k\xi}$	Start time of the unloading of commodity $O_i$ from truck $k$ under scenario $\xi \in \Xi$
$a_{i,L}^{k\xi}$	Amount of time to wait before the loading of commodity $O_i$ of truck $k$ under scenario $\xi \in \Xi$
$t_{0,L}^{k\xi}$	Departure time of service of truck $k$ from its starting depot $D_k$ under scenario $\xi \in \Xi$
$t_{n+1,U}^{k\xi}$	Arrival time of truck $k$ at its finishing depot $A_k$ under scenario $\xi \in \Xi$

The MILP model of the RSFTMDVRPTW is given as follows:

$$\text{Maximize } f_1 = \sum_{k=1}^m \sum_{i=1}^n \sum_{j=1}^{n+1} p_i x_{ij}^k \quad (1)$$

$$\text{Minimize } f_2 = TT_{\text{worst}} \quad (2)$$

Subject to:

$$\begin{aligned} & \sum_{k=1}^m \sum_{i=1}^n \sum_{j=1}^{n+1} t_{ij}^\xi x_{ij}^k \\ & + \sum_{k=1}^m \sum_{j=1}^{n+1} t_{0,j}^{k\xi} x_{0,j}^k + \sum_{k=1}^m \sum_{i=1}^n \sum_{j=1}^n t_{ij}^\xi x_{ij}^k + \sum_{k=1}^m \sum_{i=1}^n t_{i,n+1}^{k\xi} x_{i,n+1}^k \\ & + \sum_{k=1}^m \sum_{i=1}^n (w_{i,L}^{k\xi} + t_{i,U}^{k\xi} - t_{i,L}^{k\xi} - t_i^\xi) \leq TT_{\text{Worst}}, \forall \xi = 1, \dots, NS \end{aligned} \quad (3)$$

$$\sum_{j=1}^{n+1} \sum_{k=1}^m x_{ij}^k \leq 1 \quad \forall i = 1, \dots, n \quad (4)$$

$$\sum_{i=0}^n \sum_{k=1}^m x_{ij}^k \leq 1 \quad \forall j = 1, \dots, n \quad (5)$$

$$\sum_{j=1}^{n+1} x_{0j}^k = 1 \quad \forall k = 1, \dots, m \quad (6)$$

$$\sum_{i=0}^n x_{ih}^k - \sum_{j=1}^{n+1} x_{hj}^k = 0 \quad \forall h = 1, \dots, n, \quad \forall k = 1, \dots, m \quad (7)$$

$$\sum_{i=0}^n x_{i,n+1}^k = 1 \quad \forall k = 1, \dots, m \quad (8)$$

$$x_{i,0}^k = 0 \quad \forall k = 1, \dots, m, \quad \forall i = 0, \dots, n+1 \quad (9)$$

$$x_{n+1,i}^k = 0 \quad \forall k = 1, \dots, m, \quad \forall i = 0, \dots, n+1 \quad (10)$$

$$D_k^{\min} \leq t_{0,L}^{k\xi}, \quad \forall k = 1, \dots, m, \quad \forall \xi = 1, \dots, NS \quad (11)$$

$$t_{n+1,U}^{k\xi} \leq A_k^{\max} \quad \forall k = 1, \dots, m, \quad \forall \xi = 1, \dots, NS \quad (12)$$

$$L_i^{\min} \leq t_{i,L}^{k\xi} \leq L_i^{\max}, \quad \forall i = 1, \dots, n, \quad \forall k = 1, \dots, m, \quad \forall \xi = 1, \dots, NS \quad (13)$$

$$U_i^{\min} * \sum_{j=1}^{n+1} x_{ij}^k \leq t_{i,U}^{k\xi} \leq U_i^{\max}, \quad \forall i = 1, \dots, n, \quad \forall k = 1, \dots, m, \quad \forall \xi = 1, \dots, NS \quad (14)$$

$$t_{i,L}^{k\xi} + t_i^\xi \leq t_{i,U}^{k\xi}, \quad \forall i = 1, \dots, n, \quad \forall k = 1, \dots, m, \quad \forall \xi = 1, \dots, NS \quad (15)$$

$$t_{0,L}^{k\xi} + t_{0,i}^{k\xi} + w_{i,L}^{k\xi} \leq t_{i,L}^{k\xi} + M * (1 - x_{0i}^k) \quad \forall i = 1, \dots, n, \quad \forall k = 1, \dots, m, \quad \forall \xi = 1, \dots, NS \quad (16)$$

$$t_{0,L}^{k\xi} + t_{0,i}^{k\xi} + w_{i,L}^{k\xi} \geq t_{i,L}^{k\xi} - M * (1 - x_{0i}^k) \quad \forall i = 1, \dots, n, \quad \forall k = 1, \dots, m, \quad \forall \xi = 1, \dots, NS \quad (17)$$

$$t_{i,U}^{k\xi} + t_{i,j}^\xi + w_{j,L}^{k\xi} \leq t_{j,L}^{k\xi} + M * (1 - x_{ij}^k) \quad \forall i, j = 1, \dots, n, \quad \forall k = 1, \dots, m, \quad \forall \xi = 1, \dots, NS \quad (18)$$

$$t_{i,U}^{k\xi} + t_{i,j}^\xi + w_{j,L}^{k\xi} \geq t_{j,L}^{k\xi} - M * (1 - x_{ij}^k) \quad \forall i, j = 1, \dots, n, \quad \forall k = 1, \dots, m, \quad \forall \xi = 1, \dots, NS \quad (19)$$

$$t_{i,U}^{k\xi} + t_{i,n+1}^{k\xi} \leq t_{n+1,U}^{k\xi} + M * (1 - x_{i,n+1}^k) \quad \forall i = 1, \dots, n, \quad \forall k = 1, \dots, m, \quad \forall \xi = 1, \dots, NS \quad (20)$$

$$t_{i,U}^{k\xi} + t_{i,n+1}^{k\xi} \geq t_{n+1,U}^{k\xi} - M * (1 - x_{i,n+1}^k) \quad \forall i = 1, \dots, n, \quad \forall k = 1, \dots, m, \quad \forall \xi = 1, \dots, NS \quad (21)$$

$$x_{ij}^k \in \{0,1\} \quad \forall i, j = 0, \dots, n+1, \quad \forall k = 1, \dots, m \quad (22)$$

$$t_{i,L}^{k\xi} \geq 0, \quad \forall i = 0, \dots, n, \quad \forall k = 1, \dots, m, \quad \forall \xi = 1, \dots, NS \quad (23)$$

$$t_{i,U}^{k\xi} \geq 0, \quad \forall i = 1, \dots, n+1, \quad \forall k = 1, \dots, m, \quad \forall \xi = 1, \dots, NS \quad (24)$$

$$w_{i,L}^{k\xi} \geq 0, \quad \forall i = 1, \dots, n, \quad \forall k = 1, \dots, m, \quad \forall \xi = 1, \dots, NS \quad (25)$$

$$TT_{worst} \geq 0 \quad (26)$$

The problem represents a bi-objective optimization problem. The first objective function (1) seeks to maximize the total profit obtained from the selected commodities and the second objective function (2) aims to minimize the worst total operation time of all trucks over all considered scenarios.

Constraints (3) ensure that for all scenarios, the total operation time (empty travel time (first term), full travel time (second, third, and fourth terms), and waiting time before loading and unloading commodities (fifth term)) needed by all trucks does not exceed  $TT_{worst}$ . Constraints (4) and (5) imply that each collection and delivery location can be visited at most once. Constraint (6) guarantee that each truck must begin its journey from its starting depot. Constraints (7) ensure the conservation of flow; once a truck has packed an order, it must unload it at the corresponding delivery location. Constraint (8) guarantee that each truck finishes its route at the arrival depot. Constraints (9) and (10) ensure that each truck cannot return to its departure depot and cannot visit any point after its arrival depot. Inequalities (11)-(21) are used to compute the truck start time, the start time of the loading/unloading of commodities, and the time to wait at loading points in all scenarios. The time window constraints are respected using Inequalities (11)-(14). Constraints (15) require, at a commodity level, that the unloading time be greater than the sum of the loading time and the time from the commodity's collection location to its delivery location. Constraints (16) and (17) impose that loading a commodity onto a truck can only start after the truck has left its departure depot. Constraints (18) and (19) ensure that a truck can only pick up the next commodity after unloading the previous one, and displacement occurs. Constraints (20) and (21) guarantee that a truck can only unload a commodity if it can arrive at the arrival depot before the latest service end time. Finally, constraints (22)-(26) specify the appropriate values for decision variables.

#### IV. MILP-BASED LEXICOGRAPHIC APPROACH FOR THE RSFTMDVRPTW

Multi-objective optimization (MOO) problems involve optimizing more than one objective function simultaneously, which is usually in conflict, so improving one leads to worsening another. The lexicographic approach is a widely used solution method for MOO [21]. Fig. 2 depicts a general example of the lexicographic method in which the decision maker begins by ranking the objective functions in order of importance and then solves sequentially mono-objective problems starting with the most critical function and progressing to the least critical function. The lexicographic approach, similar to other methods (epsilon constraint, weighted sum, and so on), does not require any parameter configurations. Moreover, once the decision maker has prioritized one objective function over the other, it can provide a Pareto-optimal solution for the MOO problem.

## V. COMPUTATIONAL EXPERIMENTS

The computational experiments were performed using the AMPL programming language with CPLEX solver (version 12.7) on a laptop computer Intel Pentium Core i7- 4790 with 3.6 GHz and 16 GB of RAM memory.

The experiments are conducted across adapted SFTMDVRPTW instances proposed by EL Bouyahyious and Bellabdaoui [17], which are generated based on three classes  $R / C / RC$  of Solomon's VRPTW benchmark instances [22]. 'C' means that the points are clustered, 'R' indicates that the points are random, and 'RC' denotes that the points are both clustered and random.

In this study, the tests were restricted to the  $R$  problem class since it is most relevant to the FTVRP variant and the most difficult to solve. We used eight different instances from the datasets of El Bouyahyious and Bellabdaoui [17] with two different types of time windows ( $SFT1-4\_R25\_20\_2$  and  $SFT1-4\_R100\_50\_5$ ). We have adapted these instances to the RSFTMDVRPTW by adding a number of scenarios. A fixed number of arbitrary edges is selected in each scenario, and their travelling times are perturbed. Therefore, 96 new instances are generated and solved with the MILP-based lexicographic method.

Each instance is labelled as  $Ri\_n\_m\_NS\_l\_p$ , where:

- $i$  is the instance ID.
- $n$  gives the number of orders,  $n \in \{25, 50\}$ .
- $m$  gives the number of trucks,  $m \in \{2, 5\}$ .
- $NS$  denotes the number of scenarios,
- $NS \in \{10, 50, 100\}$ .
- $l$  denotes the uncertainty level,  $l \in \{50, 100\}$ . The travel time on each perturbed edge for each scenario varies on the interval  $[d(i, j), (1 + l\%)d(i, j)]$ , where  $d(.,.)$  denotes the Euclidean distance between any two points (shortest time)
- $p$  denotes the number of perturbed edges for every scenario, representing 10% or 20% of the total edges for each instance.

Table III summarizes the results performed on the 96 generated instances.  $TT_{worst}$  denotes the worst observation of the total travel time over all scenarios, and CPU represents the running time for CPLEX. For each instance, CPLEX is run for two-hour-time limits.

Table III shows the following observations:

- The proposed MILP-based lexicographic method performs well in all instances with 20 commodities where the CPLEX solver can provide optimal solutions in a relatively short time. When the number of commodities is increased to 50, CPLEX is unable to solve some instances optimally within 2 hours.
- As expected, the CPU time is significantly impacted by the number of commodities, the width of the time windows, and, in particular, the selective aspect (i.e.,

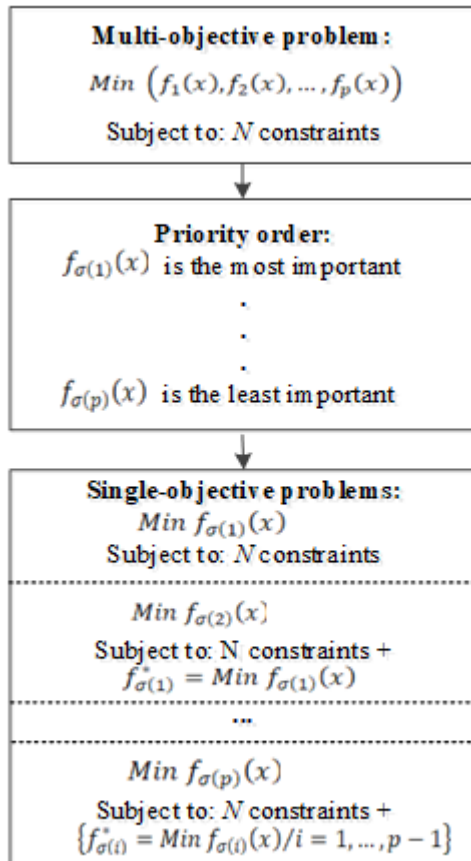


Fig. 2. Principles of the lexicographic optimization approach.

In this study, we use a lexicographic method based on the MILP formulation already defined. The selective aspect facilitates the ranking of the objective functions. This ranking was chosen on the grounds that maximizing the profit can reflect a higher quality of service, as well as on the grounds that beginning with the minimization of total travel time will generate a solution where no delivery commodity is assigned to trucks, where each route of a truck  $k$  will be the shortest path from departure point  $D_k$  to end depot  $A_k$  to obtain the lowest travel time. Therefore, we maximize the first objective  $f_1$  (the collected profit) first and then minimize the second objective  $f_2$  (the total travel time) based on the obtained solution for  $f_1$ . As a result, the original MILP model is transformed into two sequential models ( $P_1$ ) and ( $P_2$ ) as follows:

$$(P_1): f_1^* = \text{Min } f_1$$

$$\text{Subject to: Constraints (4) – (25)}$$

$$(P_2): f_2^* = \text{Min } f_2$$

$$\text{Subject to: } \begin{cases} \text{Constraints(3) – (26)} \\ f_1 \geq f_1^* \end{cases}$$

the number of unselected commodities in the obtained optimal solution). In all cases, as the number of unselected orders grows, the trucks cannot select some orders, resulting in increased CPU time.

- Furthermore, the number of scenarios and the uncertainty level directly affect the CPU time. As the values of these two parameters grow, so does the difficulty of resolving the instances.
- If the uncertainty level is set to 100, the travel time will likely be doubled. Utilizing many scenarios can diminish the uncertainty of travel time, resulting in more conservative estimates of the total worst-case travel time for all assumed scenarios.

- When comparing two different instances, a larger number of perturbed edges does not always imply a lower profit because those edges are selected randomly (e.g.,  $R1_{50_5_100_100_10}$  and  $R1_{50_5_100_100_20}$  have optimal profits of 5852 and 5841, respectively).

The robust aspect can significantly increase the problem's complexity, impacting the CPU time required to obtain a Pareto-optimal solution. Furthermore, the reported solutions are still worse than or equal to the non-robust solutions computed employing just the ideal scenario [17]. However, the feasibility of the obtained solution, over all scenarios, is the main advantage of robust optimization.

TABLE III. RESULTS OF THE PROPOSED MILP-BASED LEXICOGRAPHIC METHOD ON THE 92 GENERATED INSTANCES

Instance	Profit	TT <sub>worst</sub> *	CPU (s)	Instance	Profit	TT <sub>worst</sub> *	CPU (s)
R1_20_2_10_50_10	2760	740	99.88	R1_50_5_10_50_10	5942*	2175*	5726.6
R1_20_2_10_50_20	2743	783	152.52	R1_50_5_10_50_20	5925	2218	6277.47
R1_20_2_10_100_10	2710	867	175.88	R1_50_5_10_100_10	5892	2302	6981.85
R1_20_2_10_100_20	2750	920	206.91	R1_50_5_10_100_20	5932	2355	6984.030
R1_20_2_50_50_10	2760	814	229.75	R1_50_5_50_50_10	5939	2249	6425.61
R1_20_2_50_50_20	2730	858	292.39	R1_50_5_50_50_20	5780	2293	6535.11
R1_20_2_50_100_10	2740	907	306.93	R1_50_5_50_100_10	5734	2342	7200
R1_20_2_50_100_20	2650	959	378.85	R1_50_5_50_100_20	5816	2394	6896.57
R1_20_2_100_50_10	2710	823	490.02	R1_50_5_100_50_10	5822*	2258	7200
R1_20_2_100_50_20	2760	863	557.05	R1_50_5_100_50_20	5759	2298	6615.23
R1_20_2_100_100_10	2740	922	621.41	R1_50_5_100_100_10	5841	2357	6841.13
R1_20_2_100_100_20	2740	952	687.18	R1_50_5_100_100_20	5852	2387	6558.90
R2_20_2_10_50_10	1769	735	19.5	R2_50_5_10_50_10	5717	2180	6874.16
R2_20_2_10_50_20	1757	778	72.14	R2_50_5_10_50_20	5700*	2223	6926.8
R2_20_2_10_100_10	1753	862	95.5	R2_50_5_10_100_10	5667*	2307	6950.16
R2_20_2_10_100_20	1747	915	126.53	R2_50_5_10_100_20	5707*	2360	6981.19
R2_20_2_50_50_10	1744	809	149.37	R2_50_5_50_50_10	5717*	2254	7200
R2_20_2_50_50_20	1714	853	212.01	R2_50_5_50_50_20	5687*	2298	7200
R2_20_2_50_100_10	1749	902	226.55	R2_50_5_50_100_10	5697*	2347	6236.75
R2_20_2_50_100_20	1759	954	298.47	R2_50_5_50_100_20	5607*	2399	7200
R2_20_2_100_50_10	1759	818	409.64	R2_50_5_100_50_10	5667*	2263	7147.10
R2_20_2_100_50_20	1755	858	476.67	R2_50_5_100_50_20	5717*	2303	7200
R2_20_2_100_100_10	1739	917	541.03	R2_50_5_100_100_10	5697	2362	7155.41
R2_20_2_100_100_20	1724	947	606.8	R2_50_5_100_100_20	5667	2392	7182.44
R3_20_2_10_50_10	3020	1220	17.7	R3_50_5_10_50_10	6008*	3372*	5500.5
R3_20_2_10_50_20	3020	1263	70.34	R3_50_5_10_50_20	5991	3415	5553.14
R3_20_2_10_100_10	3020	1347	93.7	R3_50_5_10_100_10	5958	3499	5576.5
R3_20_2_10_100_20	3020	1400	124.73	R3_50_5_10_100_20	5998	3552	5607.53
R3_20_2_50_50_10	3020	1294	147.57	R3_50_5_50_50_10	6008	3446	5630.37
R3_20_2_50_50_20	3020	1338	210.21	R3_50_5_50_50_20	5978	3490	5693.01
R3_20_2_50_100_10	3020	1387	224.75	R3_50_5_50_100_10	5988	3539	5707.55
R3_20_2_50_100_20	3020	1439	296.67	R3_50_5_50_100_20	5898	3591	5779.47
R3_20_2_100_50_10	3020	1303	407.84	R3_50_5_100_50_10	5958	3455	5890.64
R3_20_2_100_50_20	3020	1343	474.87	R3_50_5_100_50_20	6008	3495	5957.67
R3_20_2_100_100_10	3020	1402	539.23	R3_50_5_100_100_10	5988	3554	6022.03
R3_20_2_100_100_20	2970	1432	605	R3_50_5_100_100_20	5988	3584	6087.8
R4_20_2_10_50_10	2819	1219	15.9	R4_50_5_10_50_10	5840	3530	5099.16
R4_20_2_10_50_20	2802	1262	68.54	R4_50_5_10_50_20	5820*	3354*	6702.54

R4_20_2_10_100_10	2769	1346	91.9	R4_50_5_10_100_10	5780*	3438*	6649.9
R4_20_2_10_100_20	2809	1399	122.93	R4_50_5_10_100_20	5830*	3491	6856.93
R4_20_2_50_50_10	2819	1293	128.77	R4_50_5_50_50_10	5770*	3385	6779.77
R4_20_2_50_50_20	2809	1337	215.41	R4_50_5_50_50_20	5740*	3429	7200
R4_20_2_50_100_10	2749	1386	310.95	R4_50_5_50_100_10	5700*	3311	7200
R4_20_2_50_100_20	2799	1438	332.87	R4_50_5_50_100_20	5700*	3478	6928.87
R4_20_2_100_50_10	2789	1302	424.04	R4_50_5_100_50_10	5720*	3394	7165.04
R4_20_2_100_50_20	2809	1342	500.07	R4_50_5_100_50_20	5710*	3434	6877.07
R4_20_2_100_100_10	2759	1401	520.43	R4_50_5_100_100_10	5720*	3576	7189.43
R4_20_2_100_100_20	2789	1431	560.2	R4_50_5_100_100_20	5700*	3523	7190.23

\* Indicates a feasible solution

## VI. CONCLUSIONS AND FUTURE RESEARCH

In this work, we have studied an essential variant of the full truck vehicle routing problem under uncertainty in transportation time, notably a robust selective full truckload multi-depot vehicle routing problem with time windows (RSFTMDVRPTW), in which a set of discrete scenarios represents uncertain travel times. We have proposed a discrete scenario-based MILP model for the RSFTMDVRPTW under the multi-objective facet, which addresses two conflicting functions: an economic component to be maximized and a component related to the worst observation of total operation time to be minimized. To solve the RSFTMDVRPTW, we have used an MILP-based lexicographic method, which maximizes the collected profit first and then minimizes the worst observation of total travel time based on the obtained solution for the first objective.

The considered approach was solved using CPLEX 12.6 and tested on 96 newly generated instances of up to 50 orders and five trucks adapted from the literature. The encouraging results demonstrate that the proposed lexicographic method provides a plausible Pareto-optimal solution for all instances with 20 commodities within an acceptable computing time. However, when the number of commodities is increased to 50, CPLEX cannot solve some instances optimally within 2 hours. Indeed, we remarked that the selective aspect, the values of the number of scenarios, and the uncertainty level strongly impact the proposed lexicographic method. Furthermore, the reported solutions are still worse than or equal to the nonrobust solutions computed employing only the ideal scenario. However, the feasibility of the solution over all scenarios is the main advantage of robust optimization.

As the problem is quite complex, only small instances can be solved optimally by CPLEX. Therefore, in future works, we will design an efficient metaheuristic algorithm to solve large instances of the problem with a large number of scenarios.

## REFERENCES

- [1] P. Toth and D. Vigo, *The Vehicle Routing Problem*. SIAM Monographs on Discrete Mathematics and Applications, Philadelphia, PA, 2002.
- [2] A. Annouch, K. Bouyahyaoui, and A. Bellabdaoui, *A literature review on the full truckload vehicle routing problems*. In: 2016 3rd International Conference on Logistics Operations Management (GOL), pp. 1–6. IEEE (2016).
- [3] M.O. Ball, B.L. Golden, A.A. Assad, and L.D. Bodin, *Planning for truck fleet size in the presence of a common-carrier option*, *Decision Sciences*, vol. 14, pp. 103-120, 1983.

- [4] X. Wang, and A.C. Regan, *Local truckload pickup and delivery with hard time window constraints*, *Transportation Research Part B: Methodological*, vol. 36, pp. 97-112, 2002.
- [5] V.M. Miori, *A multiple objective goal programming approach to the truckload routing problem*, *Journal of the Operational Research Society*, vol. 62, pp. 1524-1532, 2011.
- [6] J. Li, and W. Lu, *Full Truckload Vehicle Routing Problem with Profits*, *Journal of Traffic and Transportation Engineering (English Edition)*, vol. 1, pp. 146-152, 2014.
- [7] R. Liu, Z. Jiang, X. Liu, and F. Chen, *Task selection and routing problems in collaborative truckload transportation*, *Transportation Research Part E: Logistics and Transportation Review*, vol. 46, pp. 1071-1085, 2010.
- [8] N. Öner, and G. Kuyuzu, *Core stable coalition selection in collaborative truckload transportation procurement*, *Transportation Research Part E: Logistics and Transportation Review*, vol.154, 102447, 2021.
- [9] A. Wang, N. Ferro, R. Majewski, and C.E. Gounaris, *Mixed-integer linear optimization for full truckload pickup and delivery*, *Optimization Letters*, vol. 15, pp. 1847-1863, 2021.
- [10] J. Yang, P. Jaillet, and H. Mahmassani, *Real-Time Multivehicle Truckload Pickup and Delivery Problems*, *Transportation Science*, vol. 38, pp. 135-148, 2004.
- [11] D. Tjokroamidjojo, E. Kutanoglu, and G.D. Taylor, *Quantifying the value of advance load information in truckload trucking*, *Transportation Research Part E: Logistics and Transportation Review*, vol. 42, pp. 340-357, 2006.
- [12] H. Zolfagharinia, and M. Haughton, *The benefit of advance load information for truckload carriers*, *Transportation Research Part E: Logistics and Transportation Review*, vol. 70, pp. 34-54, 2014.
- [13] J. Li, G. Rong, and Y. Feng, *Request selection and exchange approach for carrier collaboration based on auction of a single request*, *Transportation Research Part E: Logistics and Transportation Review*, vol. 84, pp. 23-39, 2015.
- [14] A. Annouch, and A. Bellabdaoui, *An Adaptive Genetic Algorithm for a New Variant of the Gas Cylinders Open Split Delivery and Pickup with Two-dimensional Loading Constraints*, *International Journal of Advanced Computer Science and Applications*, vol. 12, pp. 607-619, 2021.
- [15] K. EL Bouyahyiouy, and A. Bellabdaoui, *A mixed-integer linear programming model for the selective full-truckload multi-depot vehicle routing problem with time windows*, *Decision Science Letters*, vol. 10, pp. 471-486, 2021.
- [16] K. EL Bouyahyiouy, and A. Bellabdaoui, *An ant colony optimization algorithm for solving the full truckload vehicle routing problem with profit*, *International Colloquium on Logistics and Supply Chain Management: Competitiveness and Innovation in Automobile and Aeronautics Industries (LOGISTIQUA)*, 7962888, pp. 142-147, 2017.
- [17] K. El Bouyahyiouy, A. Bellabdaoui, *The Selective Full Truckload Multi-depot Vehicle Routing Problem with Time Windows: Formulation and a Genetic Algorithm*, *International Journal of Supply and Operations Management*, vol. 9, pp. 299–320, 2022.

- [18] K. El Bouyahyiouy, and A. Bellabdaoui, *A Genetic-Based Algorithm for Commodity Selection and Full Truckload Vehicle Routing Problem*, Lecture Notes in Networks and Systems 669 LNNS, pp. 806-816, 2023.
- [19] K. El Bouyahyiouy, and A. Bellabdaoui, *An am-TSPTW transformation and a RTS algorithm for commodity selection and vehicle routing planning in full truckload industry*, 2022 IEEE 14th International Conference of Logistics and Supply Chain Management, LOGISTIQUA 2022.
- [20] F. Hammami, M. Rekek, and L.C. Coelho, *Exact and hybrid heuristic methods to solve the combinatorial bid construction problem with stochastic prices in truckload transportation services procurement auctions*, Transportation Research Part B: Methodological, vol. 149, pp. 204-229, 2021.
- [21] Y. Collette, and P. Siarry, *Multiobjective optimization: Principles and case studies*, Berlin: Springer, 2013.
- [22] M.M. Solomon, *Algorithms for the vehicle routing and scheduling problems with time window constraints*, Operations Research, Vol. 35, pp. 254-265, 1987.



# Design of Personalized Recommendation and Sharing Management System for Science and Technology Achievements based on WEBSOCKET Technology

Shan Zuo<sup>1</sup>, Kai Xiao<sup>2\*</sup>, Taitian Mao<sup>3</sup>

School of Public Administration, Xiangtan University, Xiangtan 411105, China<sup>1,3</sup>

Library Information Center, Changsha Aeronautical Vocational and Technical College, Changsha 410124, China<sup>1</sup>

School of Life Sciences, Central South University, Changsha 410031, China<sup>2</sup>

**Abstracts**—Scientific research is becoming more and more crucial to contemporary society as the backbone of the nation's innovation-driven development. The rapid growth of information technology and the rise of information technology in scientific research both contribute to the globalization of scientific research. Small research groups still don't have a place to showcase and share their accomplishments, though. In order to integrate scientific research information and combine personalised recommendation technology to suggest developments of interest to users through their historical behaviour data, the study proposes a personalised recommendation and sharing management system for scientific and technological achievements based on the Ruby on Rails framework. According to the testing results, the system had a 299ms request response time, a maximum 1KB request resource size, and a 20ms data transfer time. Additionally, the study's user-based collaborative filtering recommendation algorithm has an accuracy rate of 41% when the nearest neighbor parameter is set to 50, there are 10 information suggestions, and there are 0.7 training sets, which essentially satisfies the system criteria. In conclusion, the research suggested that a personalised recommendation and sharing management system for scientific and technological accomplishments can essentially satisfy the needs of small research teams to communicate and share scientific accomplishments, as well as realise the sharing of scientific achievements.

**Keywords**—Research management; personalised recommendations; WebSocket; ruby on rails; informatization

## I. INTRODUCTION

As the scope and extent of scientific research grow with the advancements in information technology and [1] technology, research has become increasingly diverse and intricate. Along with managing projects and delivering their results, researchers need to scrutinise data, design experiments and locate literature from a vast range of information [2]. In this scenario, research management systems (RMS) offer an efficient, automated and standardised alternative [3]. In order to automate and standardise research management, an RMS can aid researchers in managing information, designing experiments, analysing data, managing projects, and efficiently presenting research results [4]. Nevertheless, the current state of RMS in China is still nascent, and a shortage of academic platforms for the exchange of innovative research between teams makes it difficult to adapt research outcomes

[5].

## II. RELATED WORK

Real-time push and real-time communication are enabled by WebSocket technology, which is widely used in web applications that require a high degree of real-time, such as live chat, games, stock quotes, etc. Bisták P. proposes a new architecture for building virtual and remote laboratories using WebSocket communication technology to develop a remote control laboratory for three-tank hydraulic systems. The results showed that the remote laboratory had been visualised in 3D on the client side and was capable of comparing non-linear feedback control with dynamic feed-forward control [6]. To address the issue of exponential growth in IoT data traffic, Al-Joboury I.M. et al. proposed an IoT blockchain architecture using WebSocket communication. They found that proof-of-stake is more streamlined and advantageous for IoT applications than proof-of-work and Byzantine fault tolerance [7]. Pala Z et al. designed a network using machine learning to overcome the problem of synchronous system operation, which affects application efficiency, by analysing and transmitting data using WebSocket technology [8]. Abdelfattah A. S. et al. suggested a dependable approach using middleware and WebSockets technology to address the problem of Web service request timeouts in the mobile experience. The methodology improved the mobile experience by reducing network consumption time to seven times that of the straight cloud approach, according to the results [9]. To promote a multi-user real-time co-reading system using WebSocket technology, Chang C.T. et al. proposed a collaborative learning co-reading performance. Experiments revealed that the system significantly impacted students' learning outcomes during co-reading and that six out of seven hypotheses were supported [10].

In e-commerce, social networking, music, and other industries, CFRA, or a Personalised Recommendation Algorithm (PRA), is frequently utilised. When evaluated using independent datasets, Lim H et al.'s newly proposed Weighted Impulse Neighbourhood Regularised Three-Factor Decomposition One-Class Collaborative Filtering algorithm (CFA) demonstrated accuracy of the first 37 predicted associations of 8.495% with an enrichment factor of 4.19 compared to random guesses [11]. A recommendation framework that integrates local differential privacy (LDP)

with collaborative filtering has been developed by Bao T et al. to address the problem of dishonest server behavior or user privacy disclosure in case of failure. The results showed that the method outperformed other differential private recommendation methods [12]. In contrast to other fuzzy algorithms and traditional algorithms, Wu Y et al. proposed an interval fuzzy number-based CFRA. This algorithm is experimentally proven to be more effective and practical in sparse datasets with more users than items, and effective in improving prediction accuracy and ranking accuracy [13]. To address the issue of malicious user reviews and the issue of a small number of reviews affecting the accuracy of recommendations, Zheng G et al. proposed a CFRA with item labelling features [14]. Experiments revealed that the algorithm could successfully address the issue of cold-start data and that the interpretation of recommendation results was convincing. Zhang J et al. suggested a unique CFRA with item labelling features, and studies revealed that the proposed method was superior to existing methods [15].

The utilization of Websecurity technology and CFA is currently under investigation by various researchers from multiple perspectives; however, its practical application design in RMS is uncommon. To construct an RMS for scientific information PR, this study incorporates Websocket technology built on the Ruby on Rails framework and ronAJAX technology for CFRA. This study analyses the use of personalised recommendation (PR) algorithms based on WebSocket technology to recommend scientific and technological achievements of interest to users. The aim is to develop a management system for PR and information sharing that is objective, comprehensible, and logically structured. The conventional dissemination of scientific and technological advancements is frequently disseminated without sufficient personalized services. This research employs a customised recommendation system founded on network socket technology, which supplies individualised suggestions for scientific and technological advancements according to the users' requirements. This initiative aims to elevate users' curiosity and acceptance of scientific and technological developments, hence improving their utilization and market penetration. The research is categorised into four primary sections: analysis of pertinent research findings, core technology design and architecture, feasibility confirmation of the recommendation system, and a summary of the obtained results.

### III. DESIGN OF A PR AND SHARING MANAGEMENT SYSTEM FOR SCIENTIFIC AND TECHNOLOGICAL ACHIEVEMENTS BASED ON WEBSOCKET TECHNOLOGY

The sharing and interchange of scientific and technological advancements has become a crucial component of the growth of the scientific research area as a result of the ongoing development of data processing and artificial intelligence technologies. This part compares the benefits and drawbacks of popular recommendation algorithms (RA) in order to further choose the best RA for the research system. It also focuses on the essential technologies and architectural design of the research sharing management system.

#### A. Design of the Key Technology and Overall Architecture of the Research Sharing Management System

The main objective of the Research Sharing Management System is to enhance collaboration and information sharing among researchers, while facilitating the advancement of scientific research. It serves as a platform for storing, managing and sharing the results of scientific research. In order to help small research teams achieve result sharing and communication, the study will design the Research Sharing Management System from three aspects: system requirements, key technologies, and overall architecture. The system requirements are divided into two sections: business requirements and performance index requirements. The business requirements need to meet the purpose of research users to use information to achieve resource sharing, so the study adopts web research sharing system with simple interaction and unified interface. The system is divided into five modules: user authentication, team management, dynamic management, RA and private information management, and the specific structure is shown in Fig. 1.

The performance indicator requirements mainly meet the user experience, so the study combines the characteristics of the research system to design a system that should meet the performance requirements of strong real-time, operability, high reliability and high security during operation. After determining the system framework through requirements analysis, a number of development techniques are required to realise the submission, acquisition, storage, publication and notification of information. The first study uses Asynchronous JavaScript and XML (AJAX) asynchronous request technology to count user likes, favourites and comments, with page content rendered in HTML and CSS, dynamic display and interaction implemented by the DOM, and data exchanged between the browser and web server in JSON data exchange format. AJAX data exchange works with several technologies to load the system on demand, reducing unnecessary data transfer and thus speeding up the response time of the interface. The study then uses WebSocket technology to synchronise information between the browser and the server. The information transfer process between the browser and the web server is shown in Fig. 2.

The research then uses a recommendation system to complete the information filtering, which builds a user preference model based on user interaction data to recommend content of interest to the user. The RA essentially connects the user to information in a certain way, helping the user to discover content of interest while pushing content to the user, often in the form of friend recommendations, history, user characteristics and personal information. The final study uses Ruby on Rails framework technology to integrate AJAX technology, WebSocket technology, RA technology and other unmentioned technologies in an orderly manner, while abstracting simple and reusable design artefacts. The Ruby on Rails framework is designed with agile development ideas such as convention over configuration, chef selection and do not repeat. It is a typical Model View Controller (MVC) framework for mapping traditional input, processing and output logic. When a user submits a request via a browser, the server uses a route to locate the appropriate controller, which

then parses the user's request and interacts with the database using the model. Once the data has been retrieved, the controller provides the information to the view layer. This information is used by the view layer to create the finished web page, which is then returned to the browser as resources such as HTML, CSS and JavaScript. Fig. 3 shows the overall system architecture after integrating the core technologies to build the RMS.

As shown in Fig. 3, the study combines the business requirements and key technologies of the system based on following the design principles of the application while meeting the requirements of scalability, high stability, ease of operation and practicality, using the ActiveJob backend job module of the Ruby on Rails framework to create the PR system.

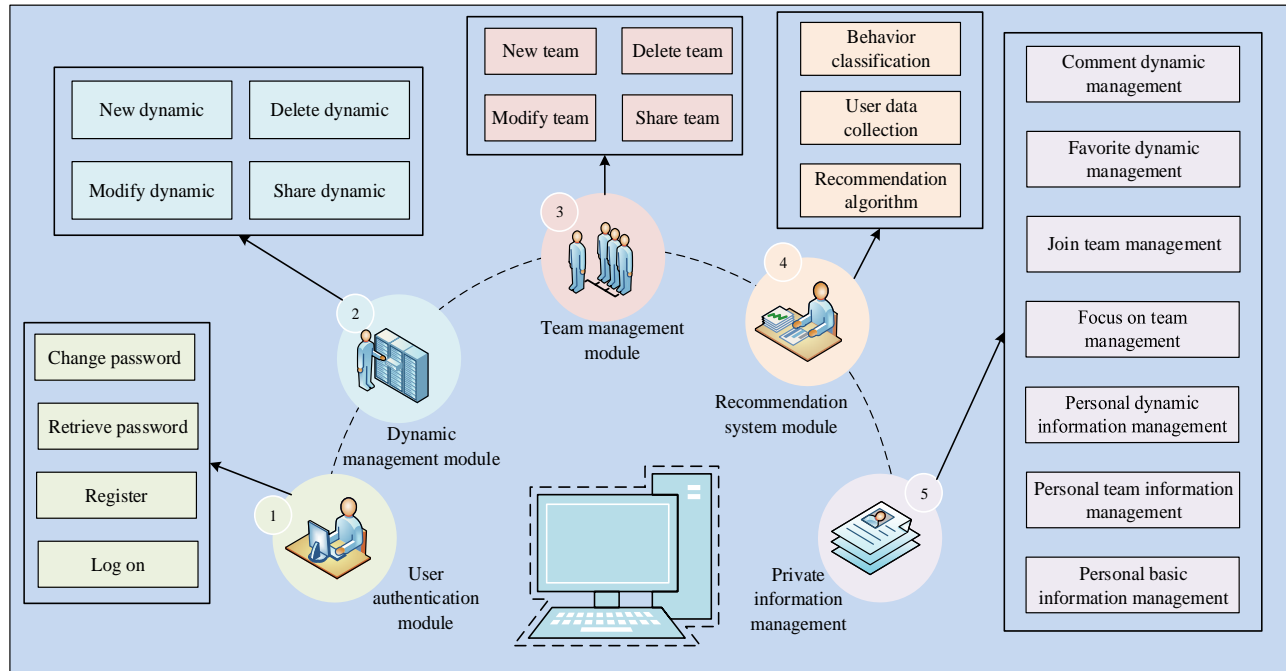


Fig. 1. Functional block diagram of scientific research sharing management platform.

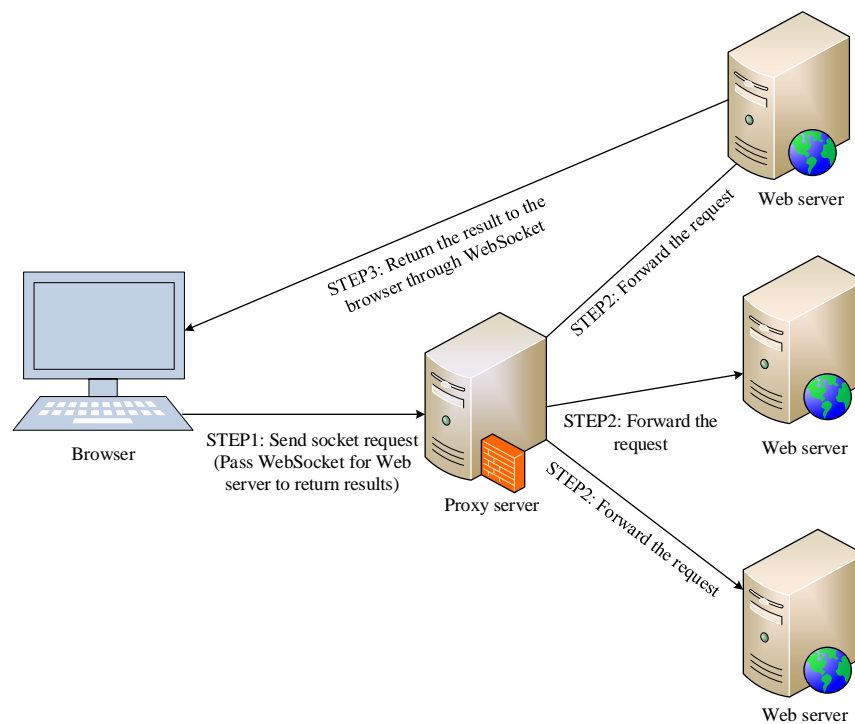


Fig. 2. Information transfer process between browser and web server.

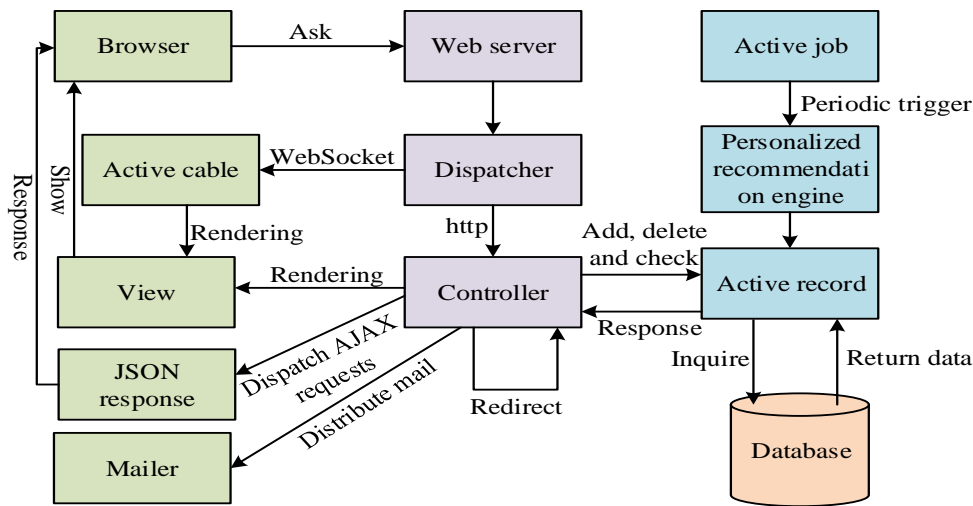


Fig. 3. System architecture design.

### B. PRA Design for Scientific and Technological Achievements based on WEBSOCKET Technology

After developing the shared management system's architecture, the study progresses to incorporate socket technology to enhance and design PRA technology accomplishments. Research users have relatively fixed research areas when doing research activities, resulting in a tendency to overlap information when searching for keywords and retrieve less available information. Although PRA serves as the foundation of PR technology, each algorithm has a unique application scenario, making it crucial to pick the best RA to suit research users' demands. Content-based algorithms, tag-based algorithms, knowledge-based algorithms, and CFRA are some examples of common RAs. The similarities between feature vectors and user preference vectors are computed by content-based RAs, which gather features from both people and items before making suggestions. Cosine similarity, as in Eq. (1) [16], is the formula used to determine similarity most frequently.

$$\cos(F_u, F_i) = \frac{F_u \cdot F_i}{\|F_u\| \times \|F_i\|} \quad (1)$$

In Eq. (1),  $F_u$  is the preference feature of a user and  $F_i$  is the preference feature of a candidate item. The closer the cosine similarity value is to 1, the closer the candidate item is to the user's preference, and the closer its value is to -1, the less suitable the candidate item is for that user. The advantages of content-based PRA are that only the user's interest features and resource attributes need to be compared online, and the similarity can be performed offline. The disadvantages are the difficulty of extracting information features from complex resources and the inability to detect similarities between similar synonyms [17]. The user's interest in the resource is determined using Eq. (2), and the tag-based RA analyzes the user's level of interest based on the number of times the user has tagged the resource. The tag-based RA then generates suggestions based on the interest matrix between the user and the resource.

$$p(u, i) = \sum_b n_{u,b} n_{b,i} \quad (2)$$

In Eq. (2),  $u$  is the user,  $i$  is the resource,  $b$  is the tag,  $n_{b,i}$  is the number of times the resource has been tagged, and  $n_{u,b}$  is the number of times the user has been tagged. The tag-based RA has the benefit of being able to comprehend the user's interests and easily obtaining the user's tags, but the drawback is that it requires work to develop the habit of tagging resources and the research user is not motivated to tag [18]. The knowledge-based RA is a system that provides a solution in response to the user's stated demands. Fig. 4 [19] illustrates the precise suggestion process.

Knowledge-based RAs are mainly applicable to specific domains and have a high recommendation accuracy rate, but the implementation process is complicated for users and not suitable for scientific users. Collaborative Filtering Architecture (CFA) is based on the user's evaluation of resources to jointly filter information and recommend content of interest to the user, which is mainly divided into user-based CFRA and item-based collaborative filtering recommendations [20]. User-based collaborative filtering uses the user's interest similarity score to make recommendations, which has no special requirements for recommended resources and can handle a variety of complex objects, while item-based CFA uses the similarity of items to make recommendations. A comparison of the advantages and disadvantages of each RA, combined with the usage scenario of the system, the study selected user-based CFRA and the specific algorithm recommendation process is shown in Fig. 5.

The complete collaborative filtering recommendation system consists of three modules: behaviour collection for collecting user information, a model for analyzing user interests, and an RA. The behaviour collection module mainly collects and classifies the operation behaviour of the user interface, and the research designs a user behaviour collection form to collect the user's research direction, likes, favourites, comments and other historical data, which is transferred to the server and then stored in the database through interface

interaction after collection. The study adopts a weighted average to infer the user's interest level, calculated as in Eq. (3).

$$\bar{p} = \frac{\sum_{i=0}^n m_i f_i(x_i)}{\sum_{i=0}^n f_i(x_i)} \quad (3)$$

In Eq. (3),  $n$  is the user action behaviour,  $m_i$  is the corresponding action behaviour weight, and  $f_i(x_i)$  is the corresponding score of the user action behaviour. The user browsing behaviour is calculated as in Eq. (4).

$$f_i(x) = \begin{cases} 1, & (\text{has viewed}) \\ 0, & (\text{has't viewed}) \end{cases} \quad (4)$$

The calculation of users' liking, favouriting and sharing behaviour is shown in Eq. (5).

$$f_i = \begin{cases} f_i(x), & (\text{thumb up}) \\ 0, & (\text{thumb down}) \end{cases}, \quad (i = 2, 3) \quad (5)$$

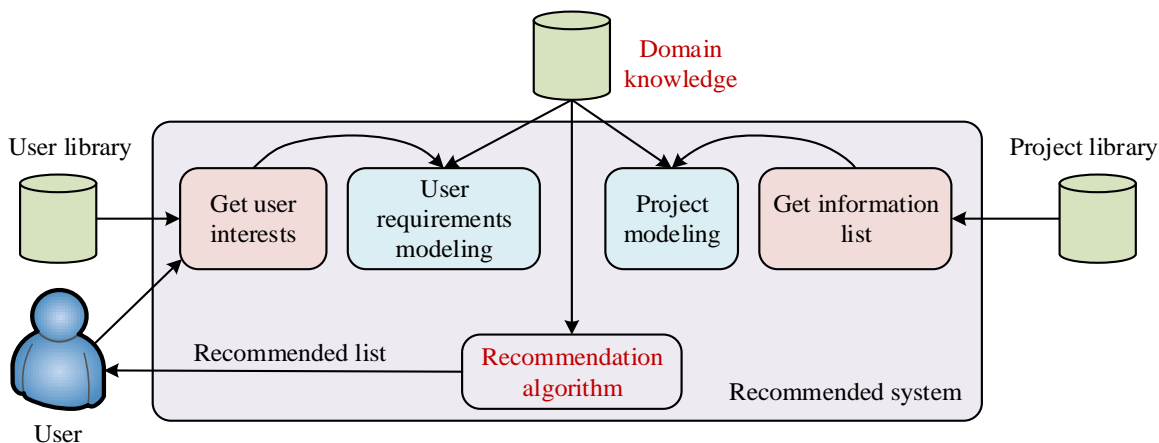


Fig. 4. Knowledge-based recommendation block diagram.

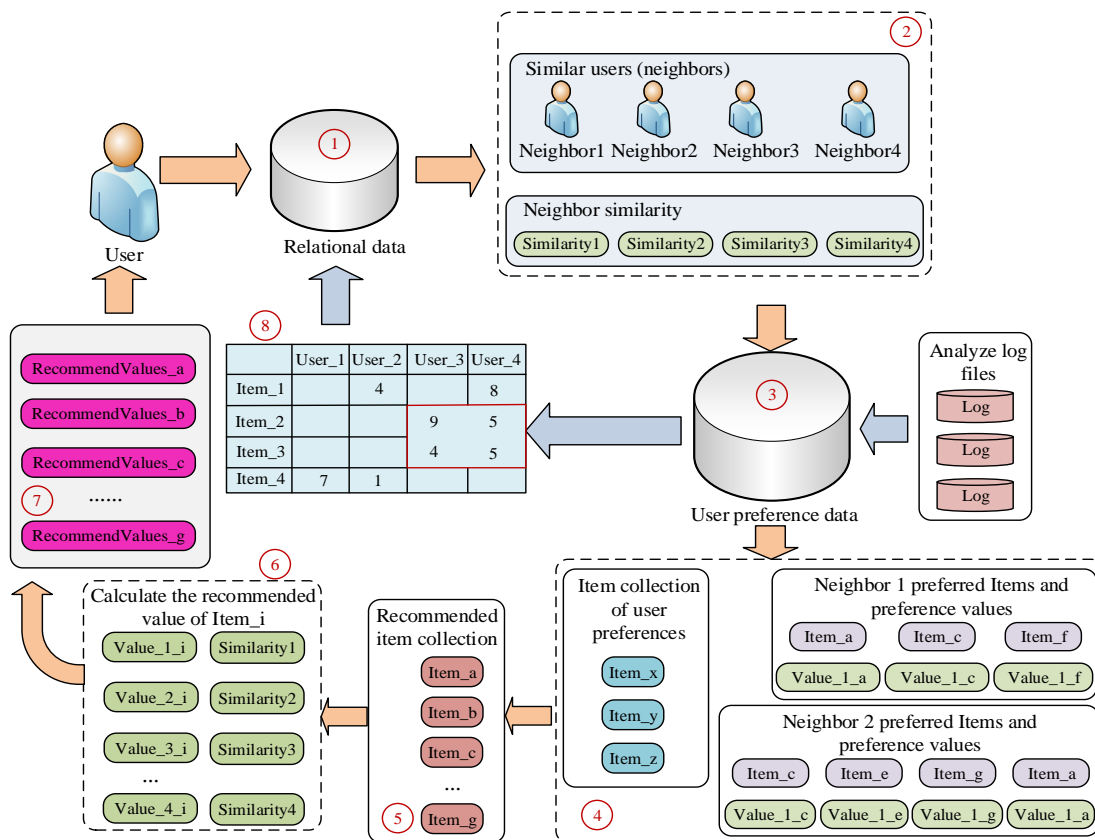


Fig. 5. CFRA structure diagram.

User objection to comment behaviour is calculated as in Eq. (6).

$$f_4(x) = x, (\text{comment statistics}) \quad (6)$$

The Euclidean distance was utilized to determine the similarity between users after the study generated the users' score matrices for each dynamic via the weighted average approach, as shown in Fig. 6. The Euclidean distance refers to  $n$  as the actual distance between two points in space.

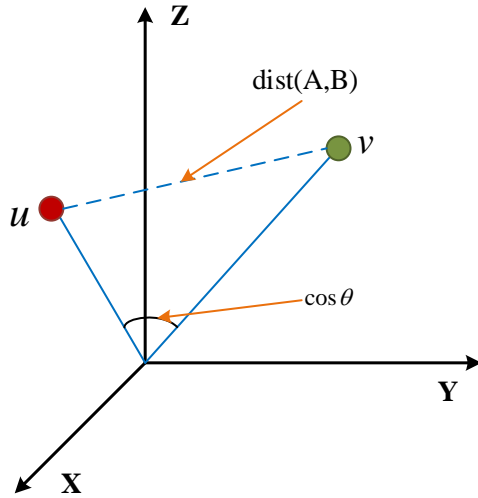


Fig. 6. Schematic diagram of the distance between two points in the three-dimensional space coordinate system.

The Euclidean distance between the user score vectors is given by Eq. (7), since the user score for each dynamic can be regarded as a multidimensional vector.

$$D(u, v) = \sqrt{\sum_{k=1}^n (x_k - y_k)^2} \quad (7)$$

The  $u$ ,  $v$  in Eq. (7) are the sets of user  $u$ ,  $v$ 's ratings of all resource actions, and  $x_i$ ,  $y_i$  are the vectors of user  $u$ ,  $v$ 's ratings of all resources respectively. The closer the Euclidean distance, the higher the similarity, and for the convenience of calculation the normalisation process is performed as in Eq. (8).

$$\begin{cases} s(u, v) = \frac{1}{1 + D(u, v)} \\ 0 \leq s(u, v) \leq 1 \end{cases} \quad (8)$$

The user with the highest similarity to the user is identified and the difference set of the rating dynamics between these two users is determined. This process is repeated until all users have been compared. The study uses accuracy, recall, coverage and Mean Absolute Error (MAE) as performance indicators to evaluate the RA. Accuracy is assessed using Eq. (9).

$$\text{Accuracy} = \frac{\sum_u |R(u) \cap T(u)|}{\sum_u T(u)} \quad (9)$$

In Eq. (9)  $R(u)$  is the set of  $N$  resources recommended to user  $u$ , and  $T(u)$  is the set of resources preferred by user  $u$  on the test set. The recall is calculated as in Eq. (10).

$$\text{Recall} = \frac{\sum_u |T(u) \cap R(u)|}{\sum_u |T(u)|} \quad (10)$$

The coverage ratio is calculated as in Eq. (11).

$$\text{Coverage} = \frac{U_{u \in U} R(u)}{|I|} \quad (11)$$

In Eq. (11),  $U$  is the set of users and  $I$  is the total number of resources.  $MAE$  is calculated as in Eq. (12).

$$MAE = \frac{\sum_{r_{u,i} \in T} |p_{u,i} - r_{u,i}|}{r_{u,i}} \quad (12)$$

In Eq. (12),  $T$  is the test set and  $r_{u,i}$  is the rating of item  $i$  by user  $u$ . In conclusion, the system RA's analysis and design are finished.

#### IV. PERFORMANCE ANALYSIS OF A PR AND SHARING MANAGEMENT SYSTEM FOR SCIENTIFIC AND TECHNOLOGICAL ACHIEVEMENTS BASED ON WEBSOCKET TECHNOLOGY

To verify the effectiveness of the designed RMS and the feasibility of the PRA, this section is designed for comparative experiments of key system technologies and tests of the accuracy and resistance to attack of the user-based CFA.

##### A. Performance Analysis of Key Technologies for Technology Sharing Management System

To verify the feasibility of the AJAX data exchange technology selected for the study, the study used AJAX and traditional HTTP on the system to respond to the same operation respectively, and the response results of both were counted. The training of the model utilises Adam as the optimizer and employs a preheating training methodology. The batch size has been established as 254, with a total iteration count of 3000 and an initial learning rate of 0.0001.

Table I shows a comparison of the results of AJAX and HTTP request responses. The comparison shows that the response time for a single HTTP request is 1613ms and the request resource size is above 40KB, while the response time for AJAX containing the request header and request body is 299ms and the request resource size is no more than 1KB. The results show that AJAX has no additional data transfer of HTML and CSS resources during page loading, which speeds up the response speed of the interface and the study. The use of AJAX data exchange instead of traditional HTTP request

interfaces is effective. To verify the efficiency of the WebSocket real-time push technology selected for the study, the study was tested at a broadband of 30MB/s, counting the throughput of the AJAX polling and WebSocket networks and the data transfer time of the HTTP and WebSocket protocols.

Fig. 7(a) shows the comparison between AJAX polling and WebSocket network throughput. Compared to AJAX polling, WebSocket data requirements are smaller for concurrent client requests below one million and minimal for concurrent requests above one million. In terms of the amount of data communicated, WebSocket has a significant advantage with better concurrency performance. Fig. 7(b) shows the data transfer time comparison between HTTP protocol and WebSocket protocol, the data transfer time in HTTP protocol is basically around 35ms, the transfer time starts to increase when the concurrent working time is 5min and then decreases to around 30 when the data transfer time is mid 10min. The data transfer time in the WebSocket protocol is around 20ms, which varies in line with the HTTP protocol, which is known to consume time each time a connection is established and released during the transfer process. The comparison between the two shows that the WebSocket protocol has a faster

transmission time and faster real-time push. The study divided the generated CiteULike-a dataset into test sets A and B based on user-document interaction records in the CiteULike document management platform, each test set includes 2500 users, 7000 papers and 100000 user-document interaction records, the study proposed the system with the traditional RMS in two the test set was tested on two test sets, using ROC and AUC as evaluation metrics, with AUC being the area below the ROC curve.

The ROC curves and AUC values for the two systems on test sets A and B are displayed in Fig. 8. The findings reveal that on test sets A and B, the suggested RMS has AUC values of 0.973 and 0.986, compared to 0.726 and 0.667 for the conventional RMS. The results show that the AUC values of the proposed RMS on test sets A and B were 0.973 and 0.986 respectively, compared to 0.726 and 0.667 for the conventional RMS. The AUC values of the proposed system increased by 34% and 48%, indicating that the improvement of key technologies has improved the accuracy of the system recommendations and made it easier for researchers to share information and communicate their results.

TABLE I. AJAX RESPONSE COMPARED TO HTTP RESPONSE

Request	HTTP			AJAX		
	Index	Application.css	Application.js	Comments	Likeables	Collects
Name	Index	Application.css	Application.js	Comments	Likeables	Collects
Status	200	200	200	200	200	200
Protocol	h2	h2	h2	h2	h2	h2
Type	Document	Stylesheet	Script	xhr	xhr	xhr
Initiator	Other	Index	Index	Jquery.min.js	Jquery.min.js	Jquery.min.js
Size	46.1KB	103KB	82KB	864B	776B	769B
Time	456 ms	834 ms	323 ms	167 ms	65 ms	67 ms
Total	1613 ms			299 ms		

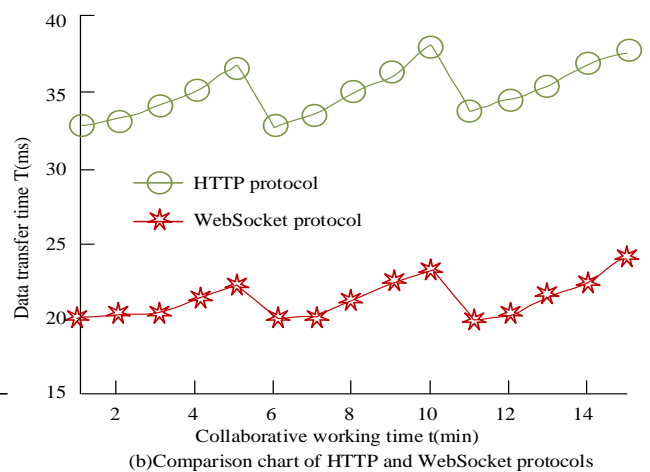
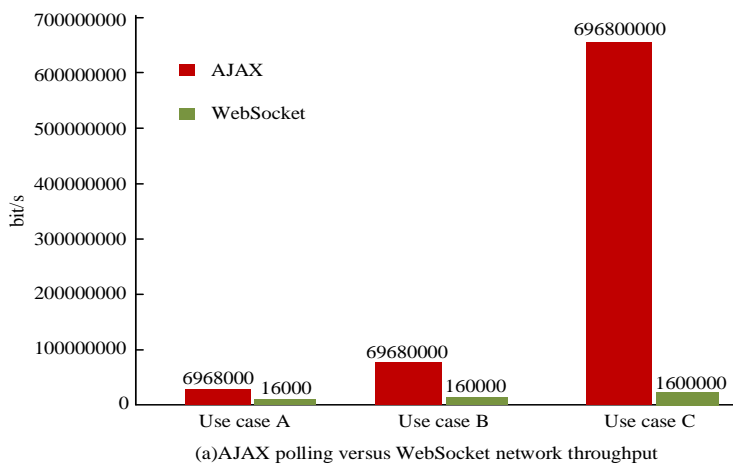


Fig. 7. AJAX polling and WebSocket network throughput and data transmission time comparison of HTTP protocol and WebSocket protocol.

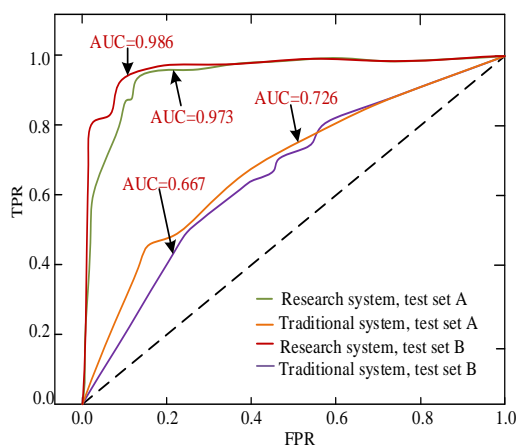


Fig. 8. ROC curves and AUC values of the two systems.

**B. Performance Analysis of the PRA for Scientific and Technical Achievements**

To verify the effectiveness of the user-based CFRA selected for the study, offline experiments were conducted on the Movielen dataset to generate TopN recommendations for each user, using accuracy, recall, coverage and popularity as performance measures, specifying the nearest neighbour parameter as the K users with the most similar interests to the recommended user, and recording the test results.

Table II shows the results of the experimental tests using user-based CFRA. The experimental results show that the accuracy of information recommendation increases with the increase of the nearest neighbour parameter K. The best recommendation is achieved when K=50. When K is constant, the accuracy of information recommendation decreases and the recall, coverage and popularity increase as the number of resources recommended to the user increases. When the training set is 0.7, the accuracy of information recommendation can basically reach 41%. In conclusion, the

accuracy of information recommendations may essentially satisfy the system requirements when K=50, the number of information suggestions is 10, and the training set is 0.7. The study included the classic CFRA based on user (CF), CFRA based on user (UserCF), content-based recommendations (CB), and knowledge-based recommendation algorithm (KR) were tested on test sets A and B to confirm the algorithms' accuracy in making recommendations. The four algorithms' recommendation accuracy was compared using the MAE as a performance evaluation metric.

Fig. 9(a) and (b) show a comparison of the recommendation accuracy of the four RAs on test sets A and B. The results show that the MAE values of UserCF on the two test sets are significantly smaller than those of CF, CB and KR algorithms, and its recommendation accuracy is the highest, with the recommendation accuracy of UserCF improving by about 13.46% compared to KR, and its recommendation accuracy improving by about 10.38% compared to CB. This shows that the research selection of UserCF algorithm can meet the needs of RMS and improve the quality of system information recommendation. To further compare the attack resistance of the four algorithms, the study added mixed attack data to the original Movielen dataset, selected K=50, with fill size of 3%, 5% and 10%, and attack size of 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9% and 10%, and tested the change of recommendation accuracy of the four algorithms as the fill size and attack size kept increasing situation.

Fig. 10(a) to (c) show a comparison of the prediction bias of the four algorithms at 3%, 5% and 10% fill size. The results show that when the fill size is the same, the prediction deviation of the four algorithms fluctuates more as the attack size increases. In conclusion, the user-based CFA used for the study is better suited for RMS, which can increase the accuracy of the system's recommendations and satisfy the demands of scientific user information sharing.

TABLE II. CFRA EXPERIMENTAL TEST RESULTS

Serial Number	Parameter			Performance			
	Neighbor Parameter	Proportion Of Training Set	Recommended Quantity	Accuracy	Recall	Coverage	Popularity
1	10	0.7	10	0.3373	0.0680	0.4094	6.7834
2	20	0.7	10	0.3766	0.0759	0.3175	6.9195
3	40	0.7	10	0.4040	0.0814	0.2395	7.0310
4	50	0.7	10	0.4083	0.0823	0.2237	7.0630
5	60	0.7	10	0.4130	0.0823	0.2095	7.0882
6	50	0.6	10	0.4623	0.0699	0.2187	6.9349
7	50	0.8	10	0.3314	0.1002	0.2325	7.1696
8	50	0.9	10	0.2119	0.1280	0.2406	7.2500
9	50	0.7	5	0.4629	0.0466	0.1691	7.1538
10	50	0.7	20	0.3468	0.1398	0.2957	6.9506
11	50	0.7	30	0.3077	0.1860	0.3505	6.8740
12	50	0.7	40	0.2790	0.2249	0.3961	6.8152
13	50	0.7	50	0.2573	0.2592	0.4233	6.7601
14	50	0.7	60	0.2396	0.2896	0.4506	6.7253



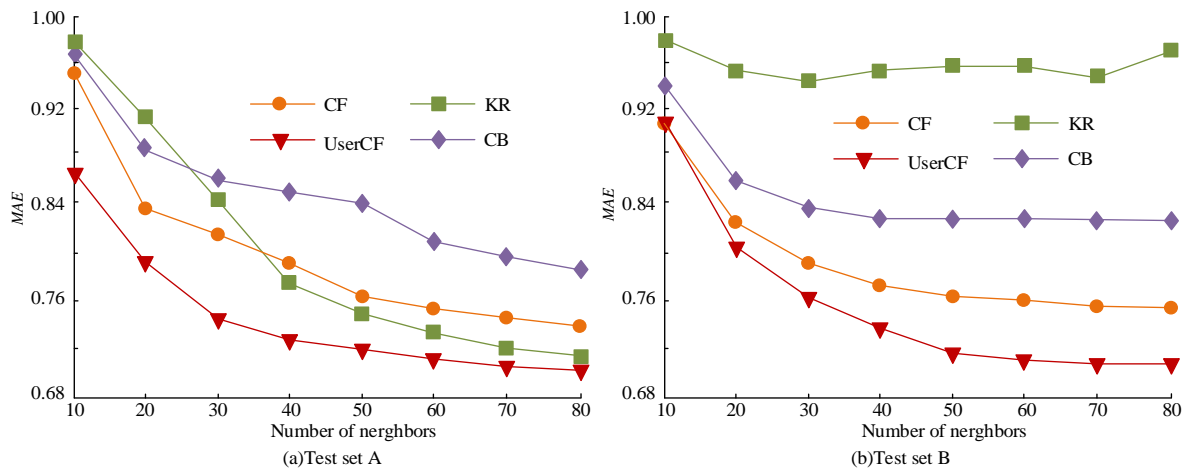


Fig. 9. Comparison of recommendation precision with different datasets.

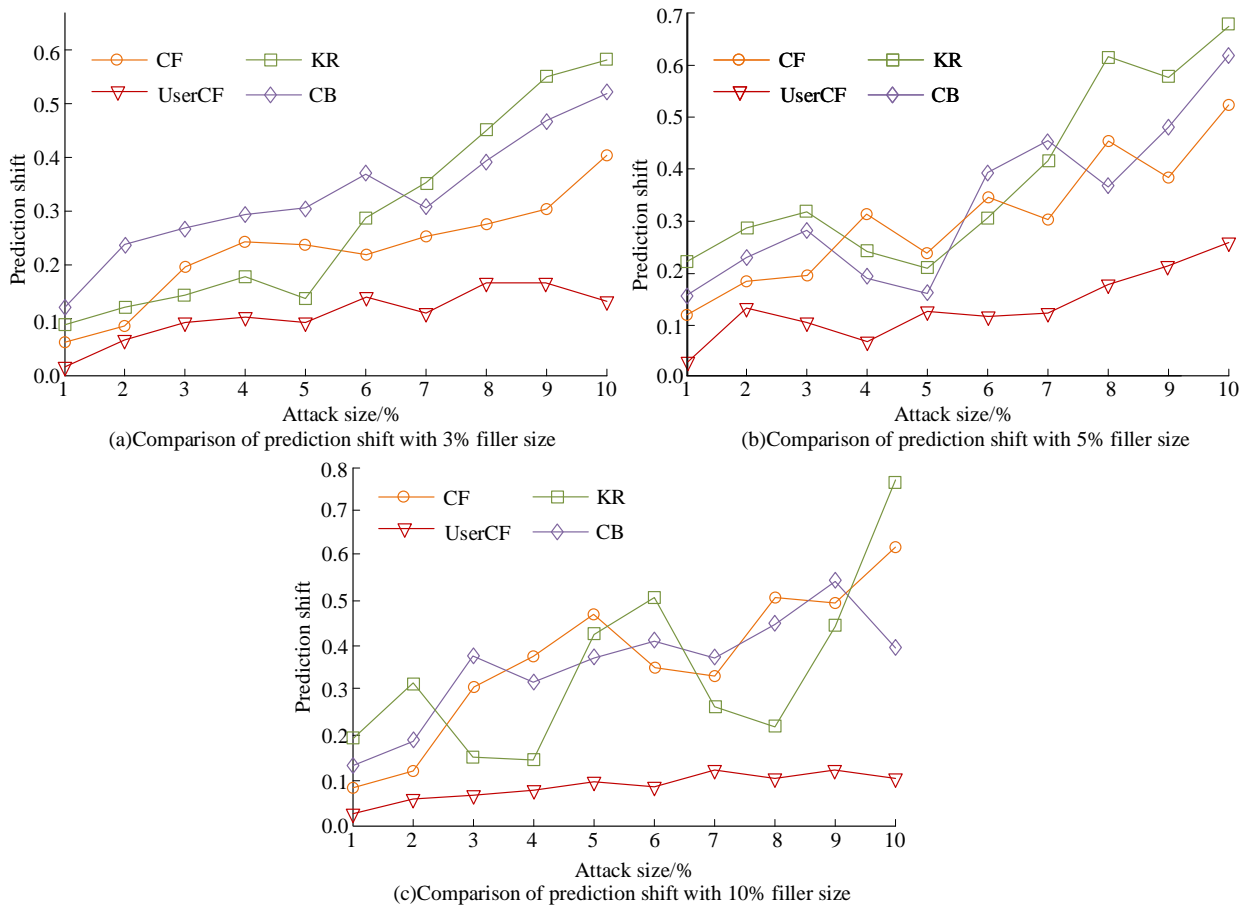


Fig. 10. Comparison of the prediction deviation of the algorithm at different filling scales.

### V. RESULTS AND DISCUSSION

With the ongoing advancements in science and technology and the rising prevalence of intelligent devices, there is a growing desire for personalized recommendations and shared management of scientific and technological accomplishments. The conventional approach to recommending and managing these accomplishments presents challenges such as information imbalance, imprecise recommendations, and burdensome administration. Designing a personalised system

for recommending and sharing scientific and technological achievements based on network socket technology has the potential to address these issues effectively, enhancing the efficiency of utilisation and improving the overall user experience. The study's experimental results demonstrated that implementing AJAX and WebSocket technologies can substantially enhance the system's response time and data transfer efficiency. Additionally, the UserCF approach proved to be more accurate in terms of information recommendation

and MAE value. The study adopted network socket technology and proposed corresponding system design and algorithm optimization based on the requirements of personalized recommendation and shared management. This holds immense importance in driving innovation and utilization of network socket technology, while also playing a guiding and exemplifying role in advancing the dissemination and collaboration of scientific and technological progress.

## VI. CONCLUSION

With the continuous development of science and technology, China has made significant progress in the field of scientific research. However, while the field of scientific research is steadily developing, the management of scientific research faces the problem of insufficient standardisation, automation and information management, so the construction and promotion of RMS is imperative. This paper proposes a scientific and technological achievement publicity and sharing management system based on WEBSOCKET technology to solve the problem of the lack of a professional platform for scientific communication and exchange among small scientific research teams. The trial results demonstrated that the system using AJAX technology has a response time that is 1314ms faster than that of traditional HTTP, that no request resource exceeds 1KB in size, and that the WebSocket technology used to transmit data demands is more efficient, with data transmission times of roughly 20ms. According to the study, the system has AUC values of 0.973 and 0.986 on the same test set, which is an improvement of 34% and 48% over conventional RMS, respectively. The UserCF method selected for the study also satisfies the system requirements for scientific research with an accuracy rate of about 41% for information recommendations at K=50, several information recommendations of 10, and a training set of 0.7. The recommendation accuracy of UserCF is around 13.46% higher compared to KR and about 10.38% higher compared to CB, which has the highest recommendation accuracy and the strongest resilience to attacks. The MAE of UserCF on the test set is much lower than that of the CF, CB and KR algorithms. In conclusion, the study claims that RMS-based systems can facilitate communication between small research teams and enable the dissemination of results. However, the study still has some shortcomings in that it collected too little information about users' personal lives, making it difficult to provide non-personalised dynamic recommendations. Future research can focus on the front-end interface of the system, user interaction, data collection and other in-depth research topics.

## FUNDINGS

The research is supported by: Research on the evaluation system of scientific and technological achievements from the perspective of science and technology finance, Supported by Hunan Provincial Innovation Foundation for Postgraduate, (No. QL20230163).

## REFERENCE

[1] Zan J. Research on robot path perception and optimization technology based on whale optimization algorithm. *Journal of Computational and Cognitive Engineering*, 2022, 1(4):201-208.

[2] Shahmirzadi T. Feasibility study of Scientific Information Visualization System of Agricultural Research, Education and Extension Organization. *Agricultural Information Sciences and Technology*, 2020, 3(5):31-40.

[3] Ding J, Wu Y, Ni X, Wang Q, Chen Y, Ye Y, Zhang X, Ma Y, Yang W. A direct coupling analysis method and its application to the Scientific Research and Demonstration Platform. *Journal of Hydrodynamics*, 2021, 33(1):13-23.

[4] Zhao L. Problems and Suggestions for the Scientific Research Management System of Universities. *Journal of Contemporary Educational Research*, 2021, 5(6):31-35.

[5] Mashizume Y, Watanabe M, Fukase Y, Zenba Y. Experiences within a cross-cultural academic exchange programme and impacts on personal and professional development. *British Journal of Occupational Therapy*, 2020, 83(12):741-751.

[6] Bisták P. Remote control laboratory for three-tank hydraulic system using matlab, websockets and javascript. *IFAC-PapersOnLine*, 2020, 53(2):17240-17245.

[7] Al-Joboury I M, Al-Hemiary E H. Consensus algorithms based blockchain of things for distributed healthcare. *Iraqi Journal of Information and communication technology*, 2020, 3(4):33-46.

[8] Pala Z, Şana M. Attackdet: Combining web data parsing and real-time analysis with machine learning. *J. Adv. Technol. Eng. Res*, 2020, 6(1):37-45.

[9] Abdelfattah A S, Abdelkader T, EI-Horbaty E I S M. RAMWS: Reliable approach using middleware and WebSockets in mobile cloud computing. *Ain Shams Engineering Journal*, 2020, 11(4):1083-1092.

[10] Chang C T, Tsai C Y, Tsai H H, Li Y J, Yu P T. An online multi-user real-time seamless co-reading system for collaborative group learning. *International Journal of Distance Education Technologies (IJDET)*, 2020, 18(4):51-70.

[11] Lim H, Xie L. A new weighted imputed neighborhood-regularized tri-factorization one-class collaborative filtering algorithm: Application to target gene prediction of transcription factors. *IEEE/ACM transactions on computational biology and bioinformatics*, 2020, 18(1):126-137.

[12] Bao T, Xu L, Zhu L, Wang L, Li R, Li T. Privacy-preserving collaborative filtering algorithm based on local differential privacy. *China Communications*, 2021, 18(11):42-60.

[13] Wu Y, ZHao Y, Wei S. Collaborative filtering recommendation algorithm based on interval-valued fuzzy numbers. *Applied Intelligence*, 2020, 50(9):2663-2675.

[14] Zheng G, Yu H, Xu W. Collaborative filtering recommendation algorithm with item label features. *International Core Journal of Engineering*, 2020, 6(1):160-170.

[15] Zhang J, Yang J, Wang L, Jiang Y, Qian P, Liu Y. A novel collaborative filtering algorithm and its application for recommendations in e-commerce. *Computer Modeling in Engineering & Sciences*, 2021, 126(3):1275-1291.

[16] Ejegwa P A, Agbetayo J M. Similarity-distance decision-making technique and its applications via intuitionistic fuzzy pairs. *Journal of Computational and Cognitive Engineering*, 2023, 2(1):68-74.

[17] Ohtomo K, Harakawa R, Ogawa T, Haseyama M, Iwahashi M. Personalized Recommendation of Tumblr Posts Using Graph Convolutional Networks with Preference-aware Multimodal Features. *ITE Transactions on Media Technology and Applications*, 2021, 9(1):54-61.

[18] Huang Y, Huang W J, Xiang X L, Yan J J. An empirical study of personalized advertising recommendation based on DBSCAN clustering of sina weibo user-generated content. *Procedia Computer Science*, 2021, 183(8):303-310.

[19] Chen X, Xue Y, Shiue Y. Rule based Semantic Reasoning for Personalized Recommendation in Indoor O2O e-commerce. *International Core Journal of Engineering*, 2020, 6(1):309-318.

[20] Zheng K, Yang X, Wang Y, Zheng X. Collaborative filtering recommendation algorithm based on variational inference. *International Journal of Crowd Science*, 2020, 4(1):31-44.

# Mechatronics Design and Development of T-EVA: Bio-Sensorized Space System for Astronaut's Upper Body Temperature Monitoring During Extravehicular Activities on the Moon and Mars

Paul Palacios<sup>1</sup>, Jose Cornejo<sup>2</sup>, Juan C. Chavez<sup>3</sup>, Carlos Cornejo<sup>4</sup>, Jorge Cornejo<sup>5</sup>, Mariela Vargas<sup>6</sup>,  
Natalia I. Vargas-Cuentas<sup>7</sup>, Avid Roman-Gonzalez<sup>8</sup>, Julio Valdivia-Silva<sup>9</sup>

Center for Space Emerging Technologies, Canada<sup>1,3</sup>

Ciencias y Tecnologías Aeroespaciales – UNPRG (CTA-UNPRG), Universidad Nacional Pedro Ruiz Gallo, Peru<sup>1,4</sup>

Universidad Nacional del Callao, Lima, Peru<sup>1,4</sup>

Universidad Tecnológica del Perú, Lima, Peru<sup>2</sup>

Space Generation Advisory Council, Vienna, Austria<sup>2</sup>

Instituto de Investigaciones en Ciencias Biomédicas, Universidad Ricardo Palma, Lima, Peru<sup>5,6</sup>

My Talent 360, GA, US<sup>3</sup>

Image Processing Research Laboratory (INTI-Lab), Universidad de Ciencias y Humanidades, Peru<sup>7</sup>

Business on Engineering and Technology S.A.C. (BE Tech), Peru<sup>8</sup>

Aerospace Sciences & Health Research Laboratory (INCAS-Lab), Universidad Nacional Tecnológica de Lima Sur, Peru<sup>8</sup>

Centro de Investigación en Bioingeniería (BIO), Universidad de Ingeniería y Tecnología – UTEC, Lima, Peru<sup>9</sup>

**Abstract**—The exploration of the universe is progressively increasing, within this inquiry, the planet Mars and the Moon remain a mystery and challenge, as well as its colonization and civilization. Thus, in the extravehicular activities (EVA) where the astronaut will be in extreme environments performing activities such as exploration, and collection of rock and soil samples for later analysis, it should be noted that when he performs these activities, he will be exposed to extreme environmental parameters such as radiation, temperature, gravity, and many other extreme conditions. Therefore, the Center of Space Emerging Technologies (C-SET) proposed a project called T-EVA, developed into the Research Line: Space Suits and Assistive Devices, and in the Research Area: Biomechatronics and Life Support Systems, with the aim of astronaut temperature monitoring during their work outside the base station, being able to know how much their body is measuring and if they are at risk of hypothermia or hyperthermia, which could cause irreparable damage. The electronic design was made for testing both in the laboratory and outside, as well as the implementation of the lycra to mount the design, resulting in a feasible prototype that can be implemented in real situations with easy access to temperature reports.

**Keywords**—Extravehicular-activities astronauts; spacesuits; body temperature; Mars; space

## I. INTRODUCTION

The journey to Mars is a major undertaking, as it is fraught with obstacles from the start of the mission to its completion, including challenges related to the atmosphere, geology, and

distance involved (Fig. 1). This has motivated both governments and private space companies to be interested in sending manned missions to Mars or the Moon, investing resources, and sending robotic missions to explore solutions to make these planets habitable and safe for humans [1]-[3]. Among these solutions are space biomedical mechatronics projects developed by the Center for Space Emerging Technologies, on which the T-EVA Project is based [4]-[9].

Extravehicular activities (EVAs) are a fundamental part of space exploration and have been a regular feature of manned missions since the earliest days of the space program. EVAs are planned activities that take astronauts outside the spacecraft or space station to perform specific tasks in space, such as repairs, maintenance, science equipment installation, or sample collection. Extravehicular activities are extremely complex and challenging due to the extreme environment of space. These activities are inherently dangerous because astronauts performing EVAs are exposed to many hazards, such as lack of gravity, extreme temperatures, cosmic radiation, and micrometeorites [15].

Successful EVAs require meticulous planning and coordination, as well as close collaboration between astronauts and the ground control team. As space exploration continues to expand to new horizons, EVAs will remain an essential part of our journey into the universe, allowing us to conduct scientific research and develop skills and technologies for future space missions, such as those planned to explore Mars and other celestial bodies [28].

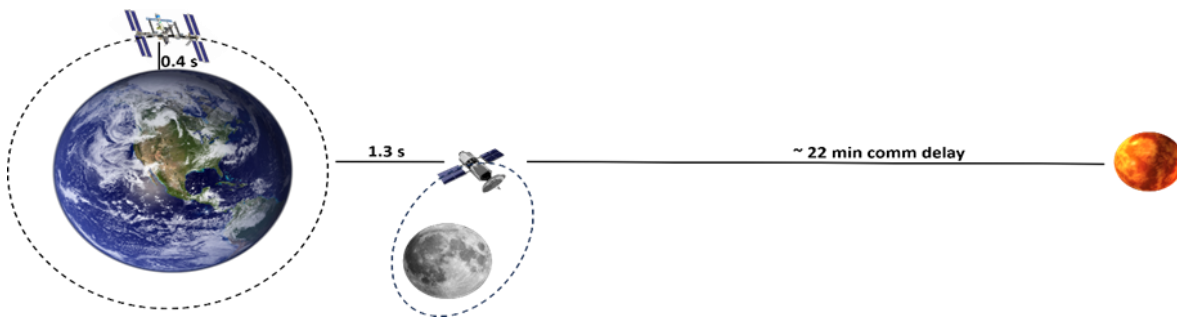


Fig. 1. Average communication delay between Earth, LEO, cislunar space, and Mars.

The lunar environment is characterized by approximately  $\frac{1}{6}$  Earth g, and surface temperatures ranging from  $-143^{\circ}\text{C}$  to  $+127^{\circ}\text{C}$  due to direct sunlight or shading in an ambient vacuum. The temperature Lunar [10], [11] seen in the image collected from Quickmap, selected Artemis 3: Candidate Landing Regions and LRD DIVINER, Polar Winter Max Temp and selected temperature range  $-173.1^{\circ}\text{C}$  to  $31.8^{\circ}\text{C}$  and LOR WAC Basemaps, WAC+NAC+NAC\_ROI\_MOSAIC.

Space studies have determined that Mars has a thin atmosphere, which does not protect the surface from dangerous

cosmic rays and micrometeorites, a problem for astronauts traveling on the surface. In addition, the presence of 95%  $\text{CO}_2$  and 0.17%  $\text{O}_2$  in the atmosphere also makes it difficult for astronauts to breathe outside their spacesuits. Extreme temperatures [12], Fig. 2(a) are also a problem for the astronauts because they range from  $-153.1^{\circ}\text{C}$  near the poles to  $19.8^{\circ}\text{C}$  near the equator [13]. The cold climate of Mars is due to the low conductivity of the surface, the sparse atmosphere, and the great distance from the sun Fig. 2(b).

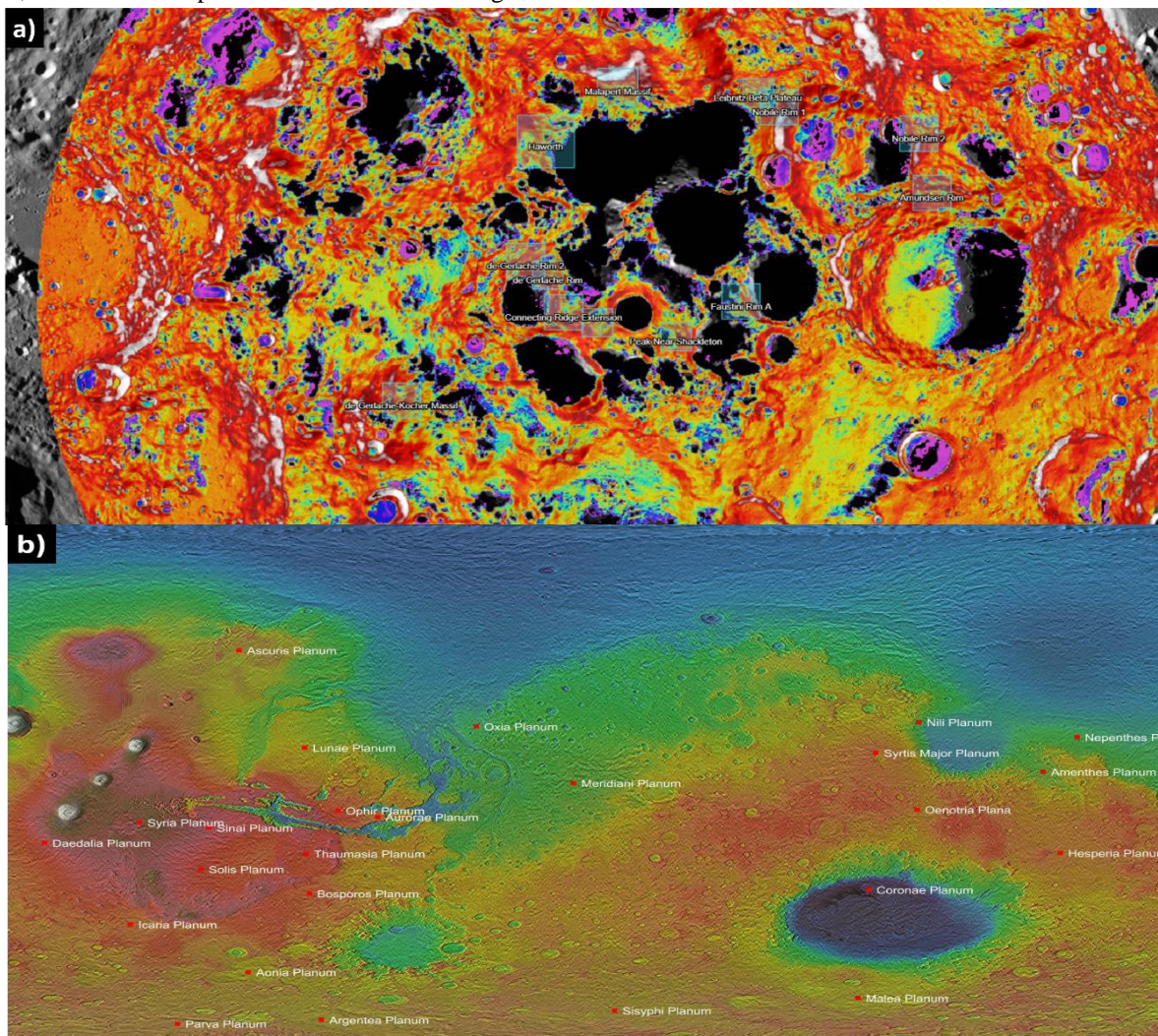


Fig. 2. (a) Temperature zones on the Moon from QUICKMAP and (b) Sea temperature from JMARS.

The human body can be conceptualized as a thermal machine that exchanges energy with its environment through moisture and heat. Likewise, thermal comfort implies a balance between the heat produced and the mechanisms of heat transfer through the effector system (vasoconstriction or vasodilation), depending on the constraints [14]. This means that exposure to adverse environments (environmental heat stress) can be detrimental to human health, especially when the environment is unknown [15]. As a result, in this temperature range in Fig. 3(a), it can be difficult for the crew to maintain the thermal

stability of both habitat and internal body temperature to conserve heat against hypothermia and its effects [16]. To solve this problem, the Extravehicular Mobility Unit (EMU) was designed to provide the necessary functions to keep the user alive [17], during extravehicular activities (EVA). As shown in Fig. 3(b), research in simulated Mars EVAs has shown that surface temperature on the suit may spatially vary by as much as 50°C depending on the astronaut's orientation relative to the sun, and atmospheric effects.

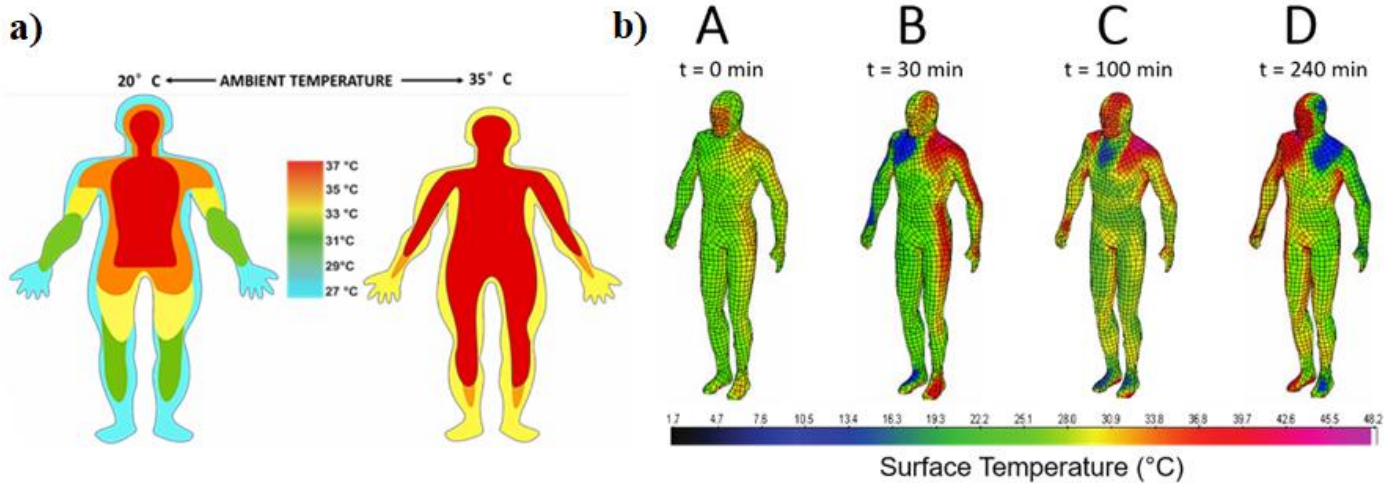


Fig. 3. (a) Thermoregulation in the human body and (b) Suit surface temperature varies throughout a simulated Martian EVA.

If the body's core temperature is exposed to extreme temperatures, it can develop hypothermia ( $<35.5\text{ }^{\circ}\text{C}$ ) or hyperthermia ( $>37.5\text{ }^{\circ}\text{C}$ ), and in peripheral parts, these effects can also occur at different degrees ( $<10^{\circ}\text{C}$ ), ( $>44^{\circ}\text{C}$ ) which can manifest in multiple symptoms [18]. Also, Table I presents a quantitative range of peripheral (skin) and core (thoracic) temperatures in relation to effects on the body and limits of thermal comfort in a thermoneutral environment [19], [20].

TABLE I. THERMOREGULATION IN HUMAN BODY

Temperature		Effects
Periphery	Body Core	
$>44-46\text{ }^{\circ}\text{C}$	$42\text{ }^{\circ}\text{C}$	Death
$36-43\text{ }^{\circ}\text{C}$	$41\text{ }^{\circ}\text{C}$	Hyperthermia
	$38-40\text{ }^{\circ}\text{C}$	Evaporation Vasodilation
$30-34\text{ }^{\circ}\text{C}$	$37\text{ }^{\circ}\text{C}$	Thermal Comfort
$24-28\text{ }^{\circ}\text{C}$	$36\text{ }^{\circ}\text{C}$	Vasoconstriction Thermogenesis
	$35\text{ }^{\circ}\text{C}$	Hypothermia
$<10\text{ }^{\circ}\text{C}$	$25\text{ }^{\circ}\text{C}$	Death

## II. PROPOSED APPLICATION FOR THE EARTH

### A. Firefighter Suit

In firefighting, activities with different levels of intensity are performed such as: throwing ladders, climbing ladders with heavy loads, performing a search, advancing a line, applying water, ventilating a roof, forcing a door, and searching a room [21]-[23]. Firefighters regularly face stress in their work, and their job performance that is directly related to saving or losing human lives, including their own [24]. The magnitude of these heat effects depends on individual factors such as age, health status, hydration, and physical fitness.

1) *Environmental conditions*: In firefighting, conducted a study compared physiological responses to an overhaul task in ambient conditions with no fire ( $15^{\circ}\text{C}$ ) to the same task performed with live fires in the structure ( $90.5^{\circ}\text{C}$  at chest level). Heart rate increased to an average of 139 bpm in the ambient conditions and to 175 beats per minute in the live-fire condition. Tympanic temperatures increased by slightly more than  $5.4^{\circ}\text{F}$  in the live-fire condition and less than  $1^{\circ}\text{F}$  in the ambient conditions as shown in Fig. 4(a), [25].

2) *Personal protective equipment (PPE)*: Protects firefighters from burn and inhalation injuries; however, due to its weight and restrictive properties, a laboratory study comparing 15 min of treadmill walking in the firefighter's uniform and 15 min of walking in fully encapsulated PPE found that the heart rate was 50 beats per minute higher while wearing the fully encapsulated gear [26].

3) *Individual characteristics*: A firefighter's age, gender, and body size all affect physiological responses to firefighting activities. In general, the risk of a heart attack while fighting a fire increase as the age of the firefighter increases.

4) *Medical condition*: High blood pressure, high cholesterol, and obesity are all well-established risk factors for cardiovascular disease. By prioritizing their cardiovascular health, firefighters can reduce their risk of experiencing cardiac events on the fire-ground and improve their overall quality of life.

5) *Fitness level*: A high level of physical fitness is necessary to successfully and safely perform demanding physical activities [27]. It increases the efficiency of the heart, improves thermal tolerance, provides cardioprotection by increasing the anticoagulant activity of the blood, and increases blood vessel dilation capacity to allow more blood to reach the muscles as shown in Fig. 4(b).

6) *Environmental control and life support subsystem*: Space is a hostile place, charged particles, solar radiation, vacuum, and free fall are potentially harmful, even fatal, to unprepared humans. Future space exploration missions will take astronauts far from Earth into extreme thermal environments, where temperature control of spacesuits will be a critical life support function. Due to wide oscillations of temperature swings, we need to balance the heat flow in, plus the heat generated internally, with the heat flow out humans, have their own, specific temperature range, where they function best. Existing thermal control technology relies on venting water to space to provide the required cooling. This approach is extremely costly, and possibly unsustainable, for future exploration missions [28].

The spacesuits have thermal regulation systems that ensure astronauts' comfort and protection by maintaining a stable internal temperature. These systems also allow the temperature of the suit to be adjusted according to external environmental conditions and the amount of heat generated by the astronauts during their metabolic activity. It has been observed that performance decrements manifest above 480 Btu/hour heat storage and tissue damage begins at 800 Btu heat storage. During the Apollo lunar surface EVAs, heat expenditure rates ranged from 780 to 1200 Btu/hour. It is important to understand and measure the estimated amount of heat expenditure prior to planetary spacewalks to ensure crew health, as the duration and requirements of the task can significantly affect heat output. Thermal management technology is an uncelebrated but nonetheless essential requirement for all spacesuits, spacecraft, and space habitats. During extravehicular activity (EVA), spacesuits must remove metabolic heat produced by the astronaut, residual heat from the suit's electronics, and absorb heat from the external environment [29]. Spacesuit design encompasses both the material selection of the spacesuit, which is important to

consider for radiation shielding and dust mitigation, as well as all the internal systems that support the regulation and monitoring of physiological health [30] as shown in Fig. 5(a).

In recent years, stretchable sensors for wearable applications have demonstrated their ability to continuously monitor health with a high level of fidelity and comfort (Table II) [31]. The integration of these elastic sensors into spacesuits could provide valuable information about astronauts' movements during EVA maneuvers, which could be combined with our proposed T-EVA to safeguard the astronaut's integrity [32]–[33] as shown in Fig. 5(b).

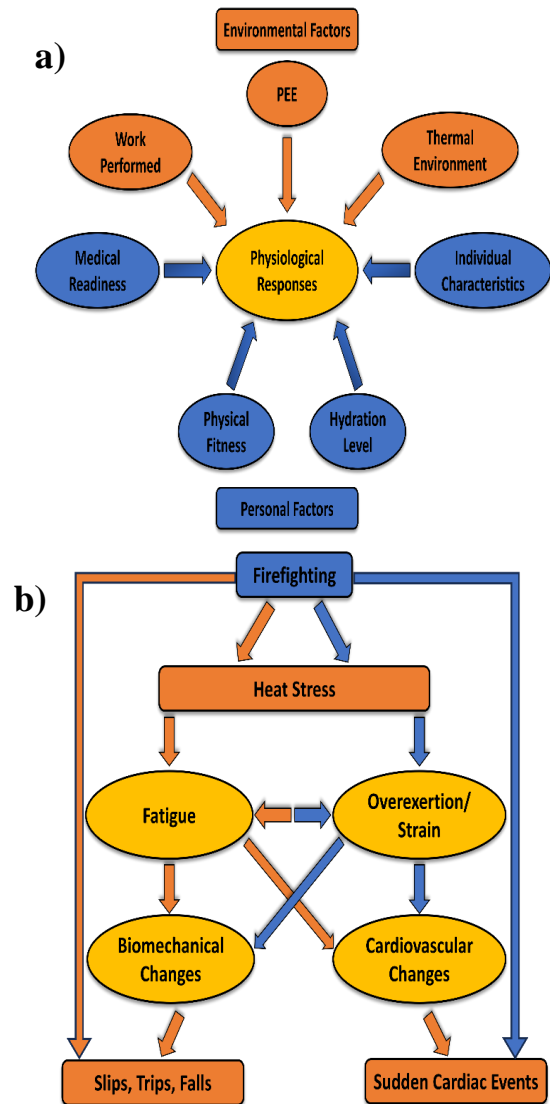


Fig. 4. (a) Factors that affect the way the body responds to firefighting activities and (b) Risk of heat stress in firefighters.

TABLE II. PLSS FUNCTIONS, RECOMMENDED TECHNOLOGIES OF CHOICE, RATIONALE, AND STRENGTHS & WEAKNESSES

Functions	Technology of Choice	Rationale	Strengths & Weaknesses
Oxygen Supply	High-Pressure Gaseous Oxygen Storage	Rechargeable on orbit, lighter & fewer parts	<b>Strengths</b> Reduced Volume Reduced System Mass Robustness Operability Reliability Less logistic <b>Weaknesses</b> Poor on-suit Mass
Thermal Control	Suit Water Membrane Evaporator	Less water contamination & can operate on Mars	
CO2 & Moisture Removal	Rapid Cycle Amine	Reduce logistics and resupply	
Power	Li-ion Polymer Batteries	Increase battery life & higher energy density	

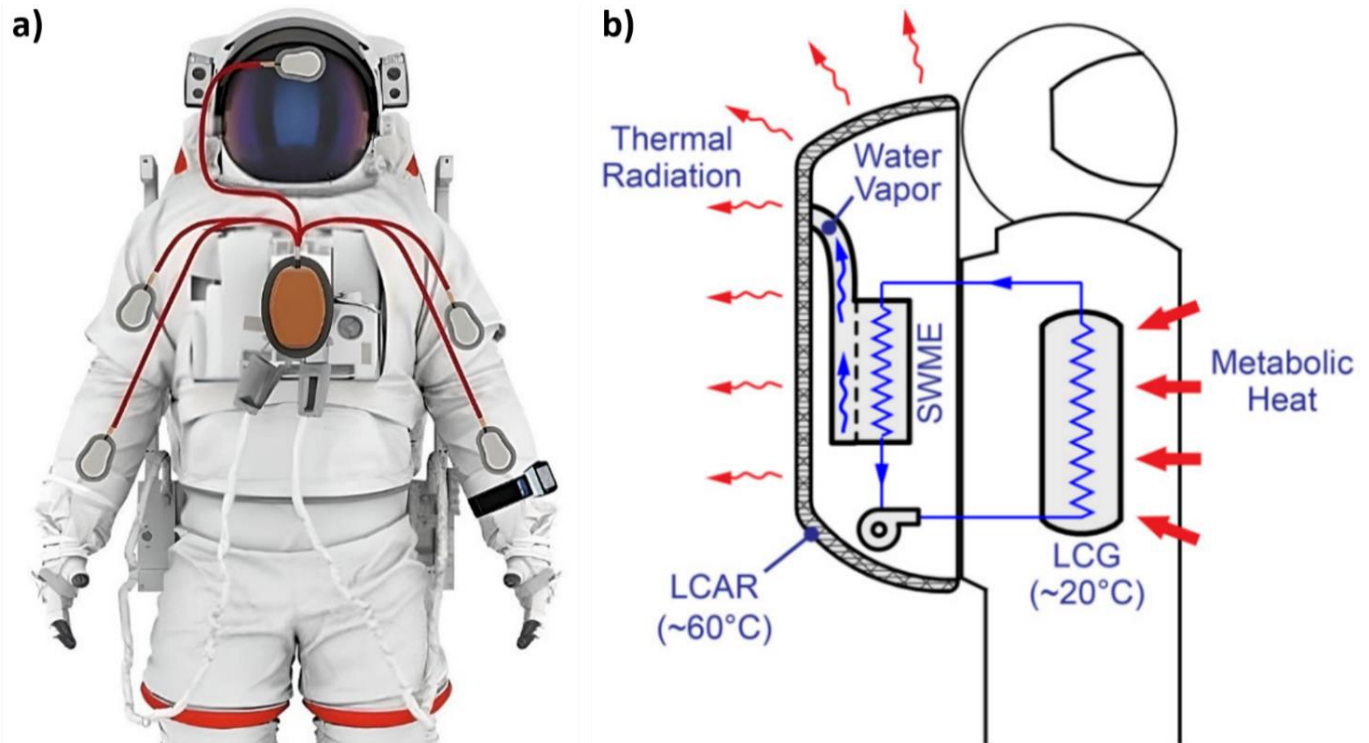


Fig. 5. (a) Spacesuit with the sensor of T-EVA and (b) The system can control the temperature inside a spacesuit without venting water.

### III. PROPOSED METHOD

This section shall describe the steps for the development of the prototype of the temperature measurement system for extravehicular activities. The proposed methodology consists of 3 phases as shown in Fig. 6.

#### A. Phase 1 / Inputs

First, as it is very well described in the methodological diagram, it is a design project focused 100% on the user. As background, one could analyze and study the success that OMEGA watches have had in the space conquest, watches that NASA astronauts have used for more than 50 years in all their space explorations. Thanks to all these experiences, one was able to define the problems of visually assisted communication and control that astronauts currently lack information assistance systems that the astronaut will need to be able to carry out high-risk extravehicular activities, and more significant challenges unknown until now. One found three immediate needs that need to be resolved, such as:

- 1) Thermoregulation

- 2) Extreme temperatures

- 3) Monitoring of physiological parameters.

#### 1) Project objectives:

a) *General Analysis:* Analysis of the context in which the project performance will be carried out, Environmental analysis, Background analysis of presented congresses and workshops, Package Evaluation, Material Evaluation, and Manufacturing evaluation.

b) *Sensor Suit:* Anthropometric Evaluation to see the anatomical shape of the sensor case. Points Location to get the temperature variation correctly.

c) *T-EVA System:* Electronic package analysis, Performance analysis, Sensor analysis, Temperature analysis, Network analysis.

d) *T-EVA Bracelet:* Electronic package analysis, Performance analysis, Visual communication analysis.

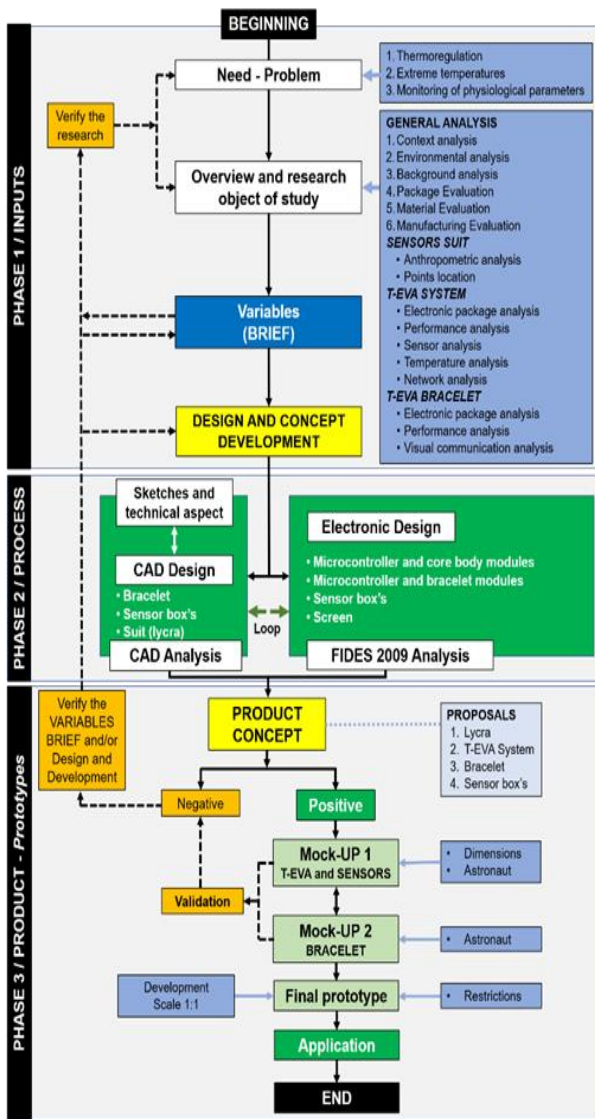


Fig. 6. Development diagram of T-EVA / I+P+p.

### B. Phase 2 / Process

In this phase, designs such as Sketches and technical aspects were made, which included analysis and CAD design with feedback from electronic design (Microcontroller and core body modules, Microcontroller and bracelet modules, Sensor boxes, Screen) [Fig. 7(a)]. This phase is divided into 3 parts.

1) *Bracelet*: The first part is a bracelet where the astronaut will be able to visualize the temperature in real-time as well as several pre-established alarms [Fig. 7(b)].

- a) Brushed titanium - Case – Grade 2.
- b) Black ceramic coated titanium -Top Ring - Grade 5.
- c) Special quartz - Glass Dome - that does not fragment.
- d) Black-coated brushed titanium-Side Push Button-Grade
- e) Titanium – Screws - Grade 2.
- f) Brushed Titanium - Bracelet Ring - Grade 2.
- g) Velcro Strap.

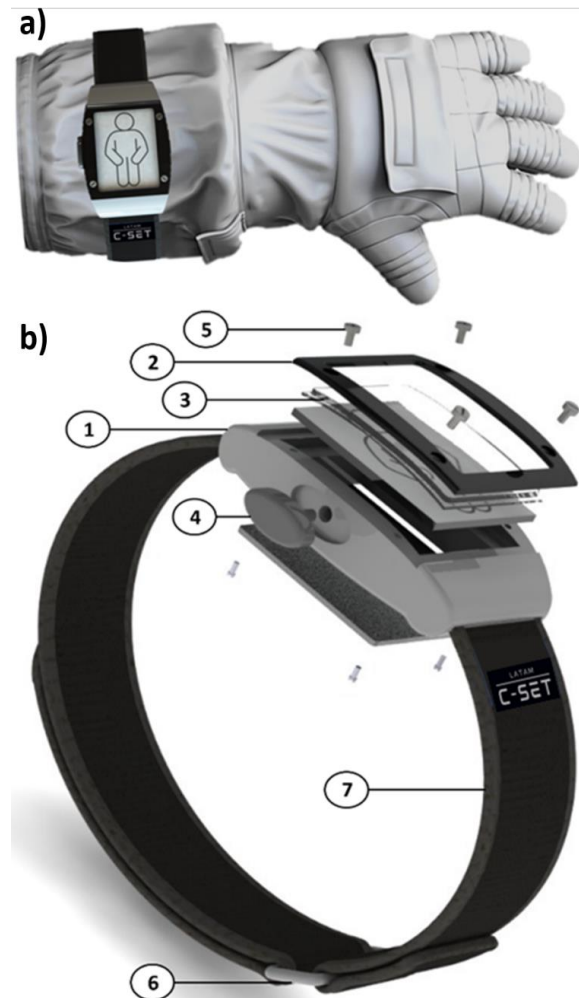


Fig. 7. (a) Bracelet in explosion view and (b) Bracelet implemented in a spacesuit.

2) *Sensors box*: The second part is the temperature sensor boxes (Fig. 8), which have an anatomical shape to capture any temperature variation, which is pretending to be manufactured based on FDA guidelines for medical devices [34], [35].

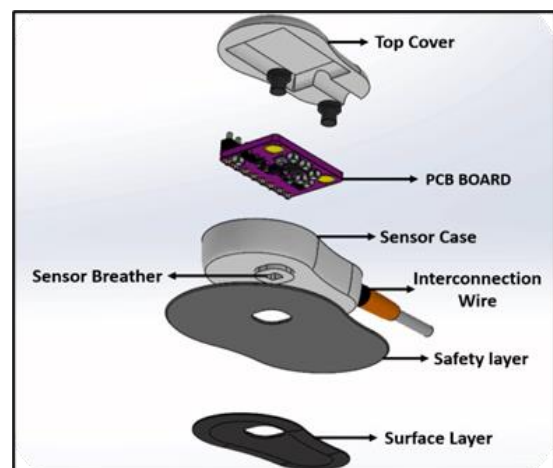


Fig. 8. Sensor Box.



3) *Central box*: The third part consists of the central box (Fig. 9), which contains the data processing board, a temperature sensor, and an RF module for sending the information to the bracelet and the base station.

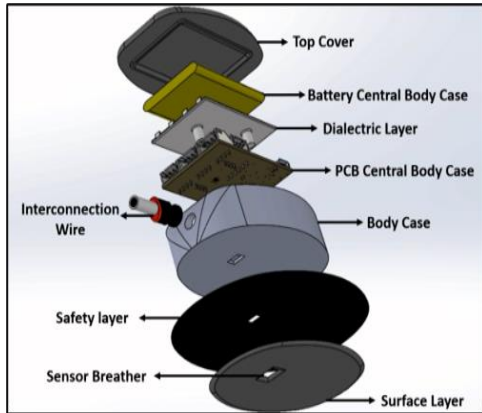


Fig. 9. Central Box.

C. Phase 3 / Product - Prototypes

1) *Implementation specifications*: Lycra being a flexible and adaptable material to the human anatomy, it is established that the measures of this would be the Latin American 95th percentile [36], Fig. 10(a) shows the location of the sensors and twisted pair wires in the lycra. In Fig. 10(b), one can see the connection. See Table III.

2) *Electronics specifications*

a) *Sensor DS18B20*: The analog temperature sensor DS18B20 [Fig. 11(a)] was used, whose voltage output is linearly proportional to temperature, generates 10mV for every 1°C, has an accuracy of  $\pm 3/4^\circ\text{C}$  in the configured range of  $-55^\circ\text{C}$  to  $150^\circ\text{C}$  [Fig. 11(b)] and has a power consumption of 60  $\mu\text{A}$ , generating a self-heating of less than  $0.1^\circ\text{C}$ , Fig. 11(c) [37].

TABLE III. SENSOR LOCALIZATION IN HUMAN BODY

SENSOR	PERIPHERY – UPPER LIMBS				BODY CORE (THORACIC)	FOREHEAD
	LEFT SIDE		RIGHT SIDE			
CODE	Upper Arm	Fore Arm	Upper Arm	Fore Arm	BC-T	FH
	L-UA	L-FA	R-UA	R-FA		

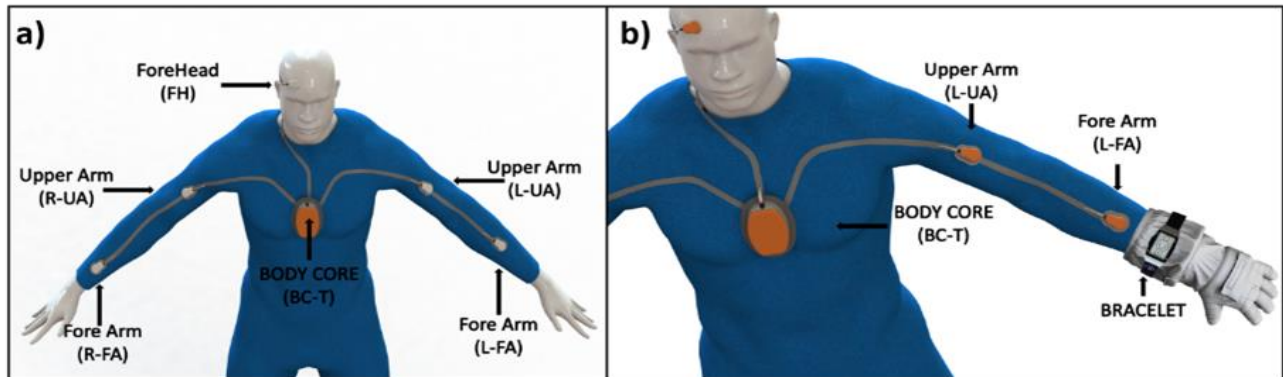


Fig. 10. (a) Sensors in the body and (b) a Bracelet mounted on the arm of the space suit.

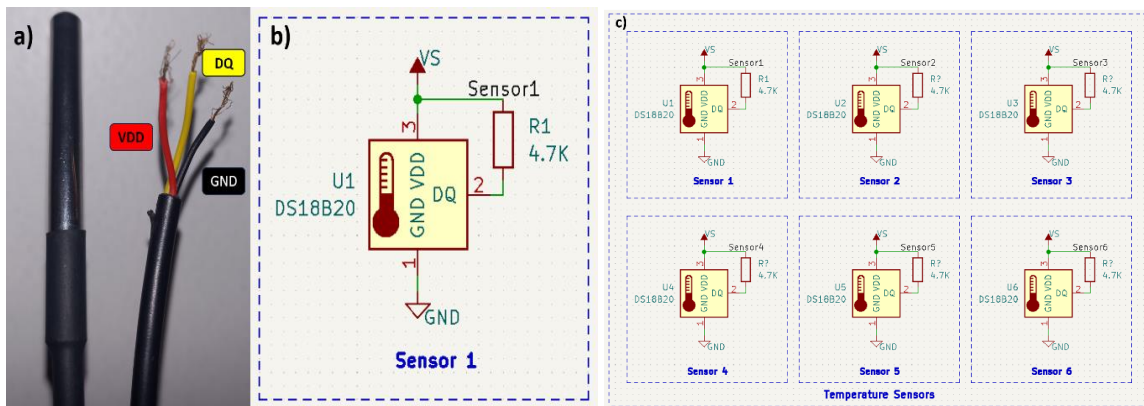


Fig. 11. (a) Sensors in the body and (b) a Bracelet mounted on the arm of the space suit.

b) *Microcontroller ESP32*: The ESP32 microcontroller was selected [Fig. 12(a), 12(b)], it has integrated wireless connectivity (WiFi and Bluetooth) to generate a communication network, has an operating temperature range of -40°C to 125°C, 30 pinouts, 512 KB RAM and the consumption specifications are shown in Table IV [38]–[41].

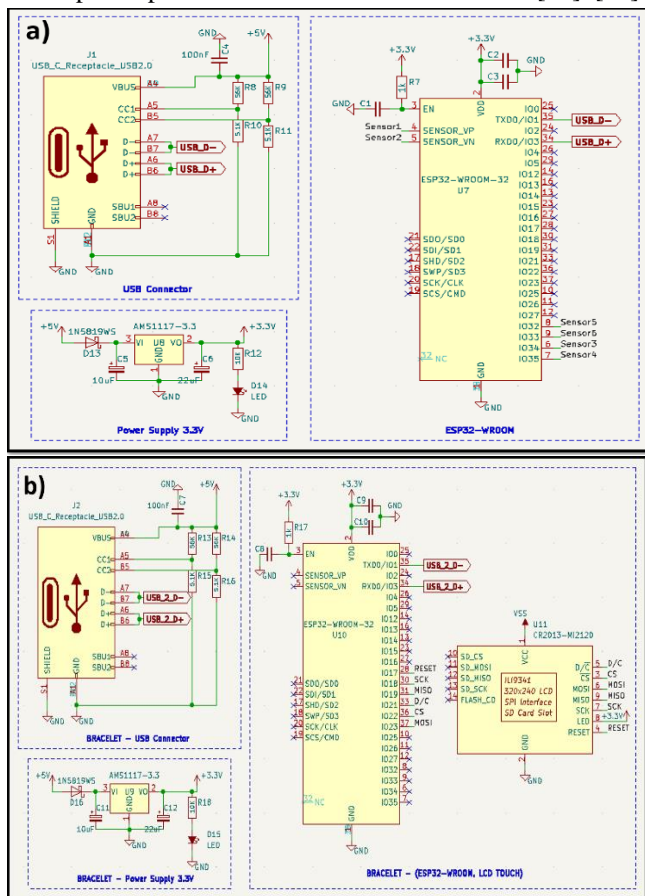


Fig. 12. (a) Schematic body core (ESP32, power supply, USB connector), and (b) Schematic bracelet (ESP32, power supply, USB connector).

TABLE IV. ELECTRONIC DESIGN OF T-EVA

Nº	Components and Consumption		
	Body Core Sensor Box	Consumption	Total
1	ESP 32 (SENDER)	180 mA	180 mA
2	DS18B20 (6)	5 mA	30 mA
3	Battery 3.7V DC – 1.4 A	Autonomy: 7.5 horas	
<b>Bracelet</b>			
1	ESP32 (RECEIVER)	80 mA	80 mA
2	TFT 2.4" – ILI9341	150 mA	150 mA
3	Battery 3.7V DC – 1A	Autonomy: 6 horas	

3) *Communication protocol*: ESP-NOW is a protocol invented by Espressif that allows connecting many devices without Wi-Fi. It is very versatile and can have unidirectional or bidirectional communication in different low-power 2.4 GHz wireless configurations. It is comparable to WiFi in the sense that pairing takes place before communication. After pairing, it becomes a secure peer-to-peer connection that does not require a handshake. This means that if one of the boards

suddenly shuts down or reboots, it will automatically connect to the other board at that time and continue to communicate. In addition, ESP-NOW can carry a payload of up to 250 bytes and can be configured to inform the application layer of the success or failure of transmission through the functions listed in Table V [42].

TABLE V. FUNCTIONS ESP-NOW

Nº	ESP-NOW PROTOCOL	
	Functions	Description
1	esp_now_init()	Wi-Fi must be initialized before initializing ESP-NOW.
2	esp_now_add_peer()	This function is used to pair a device and pass the MAC address of the peer as an argument.
3	esp_now_send()	Sends data with ESP-NOW.
4	esp_now_register_send_cb()	Registers a callback function that is triggered when sending data. This function returns whether the delivery was successful or not.
5	esp_now_register_rcv_cb()	Registers a callback function that is triggered when data is received. A specific function is called when data is received.

The configuration of an ESP32 "RECEIVER" (R) microcontroller receiving data from an ESP32 "SENDER" (S) microcontroller has been used. The communication network is unidirectional, which means that the information flows only from the sender to the receiver.

With this configuration, it is possible to collect the data from the four temperature sensors (1\_L-FA, 2\_L-UA, 3\_R-FA, and 4\_R-UA) from the sender ESP32 microcontroller and send it wirelessly to the receiver ESP32 microcontroller. After receiving the four temperature readings from the sending ESP32 microcontroller, the receiving ESP32 microcontroller displays the values on a 2.4" LCD display, as shown in Fig. 13.

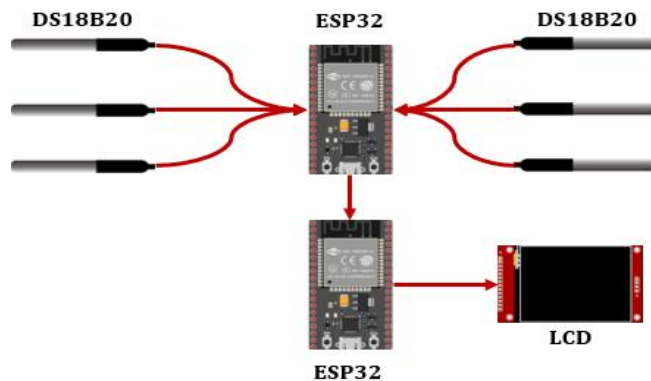


Fig. 13. Diagram of connection.

#### IV. RESULTS AND DISCUSSIONS

Once the connections and operation of the sensors and microcontroller were verified, tests were performed on a breadboard to check the code and then the system was mounted on a lycra where first one sensor was inserted and then 4 sensors (Fig. 14 and 15).



a) *First Test:* Fig. 17(a), (b) shows the sending and receiving of the 4 temperature values using the ESP-NOW protocol and the Arduino IDE serial monitor. It is observed that the data (1\_L-FA = 35.28°C, 2\_L-UA = 35.75°C, 3\_R-FA = 35.37°C and 4\_R-UA = 35.67°C) reach their destination without loss of information.

b) *Second Test:* Fig. 18 shows the sending and receiving of the 4 temperature values using the ESP-NOW protocol and the Arduino IDE serial monitor. It is observed that the data (1\_L-FA = 35.35°C, 2\_L-UA = 35.77°C, 3\_R-FA = 35.41°C and 4\_R-UA = 35.68°C) reach their destination without loss of information.

c) *Third Test:* Fig. 19 shows the sending and receiving of the 4 temperature values using the ESP-NOW protocol and the Arduino IDE serial monitor. It is observed that the data (1\_L-FA = 35.39°C, 2\_L-UA = 35.79°C, 3\_R-FA = 35.44°C and 4\_R-UA = 35.73°C) reach their destination without loss of information.

d) *Fourth Test:* Fig. 20 shows the sending and receiving of the 4 temperature values using the ESP-NOW protocol and the Arduino IDE serial monitor. It is observed that the data (1\_L-FA = 35.45°C, 2\_L-UA = 35.76°C, 3\_R-FA = 35.46°C and 4\_R-UA = 35.72°C) reach their destination without loss of information.

```
a) 22:06:03.642 -> Last Packet Send Status:      Delivery Success
22:06:04.654 -> Sent with success
22:06:04.654 -> Temperature 1_L-FA:
22:06:04.654 -> 35.28
22:06:04.654 -> Temperature 2_L-UA:
22:06:04.654 -> 35.75
22:06:04.654 -> Temperature 3_R-FA:
22:06:04.654 -> 35.37
22:06:04.654 -> Temperature 4_R-UA:
22:06:04.654 -> 35.67

b) 22:06:04.654 -> Data received:
22:06:04.654 -> Temperature 1_L-FA:
22:06:04.654 -> 35.28
22:06:04.654 -> Temperature 2_L-UA:
22:06:04.654 -> 35.75
22:06:04.654 -> Temperature 3_R-FA:
22:06:04.654 -> 35.37
22:06:04.654 -> Temperature 4_R-UA:
22:06:04.654 -> 35.67
```

Fig. 17. (a) Transmitter - Temperature data from the DS18B20 and (b) Receiver - Temperature data from the DS18B20.

```
a) 22:18:41.227 -> Last Packet Send Status:      Delivery Success
22:18:42.215 -> Sent with success
22:18:42.215 -> Temperature 1_L-FA:
22:18:42.215 -> 35.35
22:18:42.215 -> Temperature 2_L-UA:
22:18:42.215 -> 35.77
22:18:42.215 -> Temperature 3_R-FA:
22:18:42.215 -> 35.41
22:18:42.215 -> Temperature 4_R-UA:
22:18:42.215 -> 35.68

b) 22:18:42.274 -> Data received:
22:18:42.274 -> Temperature 1_L-FA:
22:18:42.274 -> 35.35
22:18:42.274 -> Temperature 2_L-UA:
22:18:42.274 -> 35.77
22:18:42.274 -> Temperature 3_R-FA:
22:18:42.274 -> 35.41
22:18:42.274 -> Temperature 4_R-UA:
22:18:42.274 -> 35.68
```

Fig. 18. (a) Transmitter - Temperature data from the DS18B20 and (b) Second test of receiving temperature data from the DS18B20.

```
a) 22:24:22.766 -> Last Packet Send Status:      Delivery Success
22:24:23.732 -> Sent with success
22:24:23.732 -> Temperature 1_L-FA:
22:24:23.732 -> 35.39
22:24:23.775 -> Temperature 2_L-UA:
22:24:23.775 -> 35.79
22:24:23.775 -> Temperature 3_R-FA:
22:24:23.775 -> 35.44
22:24:23.775 -> Temperature 4_R-UA:
22:24:23.775 -> 35.73

b) 22:24:23.731 -> Data received:
22:24:23.731 -> Temperature 1_L-FA:
22:24:23.775 -> 35.39
22:24:23.775 -> Temperature 2_L-UA:
22:24:23.775 -> 35.79
22:24:23.775 -> Temperature 3_R-FA:
22:24:23.775 -> 35.44
22:24:23.775 -> Temperature 4_R-UA:
22:24:23.775 -> 35.73
```

Fig. 19. (a) Transmitter - Temperature data from the DS18B20 and (b) Receiver - Temperature data from the DS18B20.

```
a) 22:30:45.975 -> Last Packet Send Status:      Delivery Success
22:30:46.978 -> Sent with success
22:30:46.978 -> Temperature 1_L-FA:
22:30:46.978 -> 35.45
22:30:46.978 -> Temperature 2_L-UA:
22:30:46.978 -> 35.76
22:30:46.978 -> Temperature 3_R-FA:
22:30:46.978 -> 35.46
22:30:46.978 -> Temperature 4_R-UA:
22:30:46.978 -> 35.72

b) 22:30:46.978 -> Data received:
22:30:46.978 -> Temperature 1_L-FA:
22:30:46.978 -> 35.45
22:30:46.978 -> Temperature 2_L-UA:
22:30:46.978 -> 35.76
22:30:46.978 -> Temperature 3_R-FA:
22:30:46.978 -> 35.46
22:30:46.978 -> Temperature 4_R-UA:
22:30:46.978 -> 35.72
```

Fig. 20. (a) Transmitter - Temperature data from the DS18B20 and (b) Receiver - Temperature data from the DS18B20.

## V. CONCLUSION AND FURTHER WORK

The results of this study demonstrate the feasibility of designing and implementing a prototype to measure astronaut body temperature during extravehicular activities (EVA). Body temperature ranges remained stable under normal conditions, and although problems arose with the LM35 temperature sensors, the choice of the DS18B20 sensors proved to be more successful, providing more stable and reliable readings. These sensors feature encapsulated probes that are ideal for skin contact and are water resistant, which increases their robustness when astronauts sweat. Constant temperature monitoring translates into easy-to-read reports, which is essential for preserving astronaut health during EVAs.

As a part of future work, it is intended to carry out further tests, in addition to those already performed, and to optimize the 3D printed prototype. These tests will be conducted at the Mars Desert Research Station (MDRS) in the deserts of Utah, USA, and in environments with lunar-like conditions.

The Center for Space Emerging Technologies (C-SET) is known for pursuing dual applications in every project. In this case, the T-EVA device could be employed on Earth to monitor the upper body temperature of firefighters when they face extreme heat exposure in urban or rural environments. This would provide them with real-time readings of their temperature, which would be crucial to take safety measures and protect their life and health in these challenging situations.

## ACKNOWLEDGMENT

The research has been managed, supervised by the Center for Space Emerging Technologies ([https://linktr.ee/cset\\_space](https://linktr.ee/cset_space)). Also, special thanks to the Institute of Electrical and Electronics Engineers – IEEE and to the American Society of Mechanical Engineers – ASME.

## REFERENCES

- [1] J. Rohrig, A. Himmelmann, M. Torralba, G. Quinn, P. Lee, S. R. Dalal, M. Arveng, M. Tamuly and J. Lysberg. "Development and Test of a Spacesuit Informatics System for Moon, Mars, and Further Deep-Space Exploration". 51st International Conference on Environmental Systems, 10-14 July 2022, St. Paul, Minnesota.
- [2] M. R. Islam, F. H. Chowdhury, S. Rezwani, M. J. Ishaque, J. U. Akanda, A. S. Tuhel and B. B. Riddhe. "Novel design and performance analysis of a Mars exploration robot: Mars rover mongol pothik". In 2017 Third International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN) (pp. 132-136). November, 2017.
- [3] A. Aravindhan, G. Laxmikanth and S. Kamalraj. "Medical Diagnosis During Space Tourism And Future Mars Colonization". In 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS) (pp. 557-559). IEEE, March, 2020.
- [4] J. Cornejo, J. A. Cornejo-Aguilar and R. Palomares. "Biomedik surgeon: surgical robotic system for training and simulation by medical students in Peru". In 2019 International Conference on Control of Dynamical and Aerospace Systems (XPOTRON) (pp. 1-4). IEEE, April, 2019.
- [5] J. Cornejo, J. P. Perales-Villaruel, R. Sebastian and J. A. Cornejo-Aguilar. "Conceptual design of space biosurgeon for robotic surgery and aerospace medicine". In 2020 IEEE ANDESCON (pp. 1-6). IEEE, October, 2020. doi: 10.1109/ANDESCON50619.2020.9272122.
- [6] P. Palacios, J. Cornejo, M. V. Rivera, J. L. Napán, W. Castillo, V. Tiellacuri, ... and J. C. Chávez. "Biomechatronic embedded system design of sensorized glove with soft robotic hand exoskeleton used for rover rescue missions on mars". In 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS) (pp. 1-10). IEEE, April, 2021. doi: 10.1109/IEMTRONICS52119.2021.9422634.
- [7] V. Tiellacuri, G. J. Lino, A. B. Diaz, and J. Cornejo. "Design of wearable soft robotic system for muscle stimulation applied in lower limbs during lunar colonization". In 2020 IEEE XXVII International Conference on Electronics, Electrical Engineering and Computing (INTERCON) (pp. 1-4). IEEE, September, 2020. doi: 10.1109/INTERCON50315.2020.9220206.
- [8] J. Cornejo, J. A. Cornejo-Aguilar, R. Sebastian, P. Perales, C. Gonzalez, M. Vargas, and E. F. Elli. "Mechanical design of a novel surgical laparoscopic simulator for telemedicine assistance and physician training during aerospace applications". In 2021 IEEE 3rd Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS) (pp. 53-56). IEEE, May, 2021. doi: 10.1109/ECBIOS51820.2021.9510753.
- [9] V. Tiellacuri, J. Cornejo, N. Castrejon, A. B. Diaz, K. Hinostroza, D. Dev, ... and A. Roman-Gonzalez. "Design of Biomedical Soft Robotic Device for Lower Limbs Mechanical Muscle Rehabilitation and Electrochemical Monitoring under Reduced-Gravity Space Environment". In 2021 IEEE URUCON (pp. 227-231). IEEE, November, 2021. doi: 10.1109/URUCON53396.2021.9647197.
- [10] J. P. Williams, B. T. Greenhagen, D. A. Paige, N. Schorghofer, E. Sefton-Nash, P. O. Hayne, ... and K. M. Aye. "Seasonal polar temperatures on the Moon". *Journal of Geophysical Research: Planets*, 124(10), 2505-2521, 2019.
- [11] T. M. Baugh, and R. G. Osborn. "QuickMAP". *Environmental Software*, 6(2), 115., 1990.
- [12] Java Mission-planning and Analysis for Remote Sensing (JMARS). "Welcome to the JMARS website". *Asu.edu*. [Online], 2023. Available: <https://jmars.mars.asu.edu/>. [Accessed: 24-Apr-2023].
- [13] M. Biswal, M. Kumar, D. Basanta, N. Annavarapu and R. Naidu. "Orbital and Planetary Challenges for Human Mars Exploration". arXiv e-prints, arXiv:2101. [Online], 2021 Available: <https://arxiv.org/ftp/arxiv/papers/2101/2101.04725.pdf>. [Accessed: 01-Nov-2022].
- [14] H. C. Gunga. "Human physiology in extreme environments". Academic Press., 2021.
- [15] S. P. Chappell, J. R. Norcross, A. F. Abercromby, O. S. Bekdash, E. A. Benson, S. L. Jarvis, ... and J. A. Tuxhorn. "Evidence report: risk of injury and compromised performance due to EVA operations (No. JSC-CN-39092)", 2017. Available: <https://humanresearchroadmap.nasa.gov/Evidence/reports/EVA.pdf>. [Accessed: 25-Oct-2022].
- [16] M. F. Uth, J. Koch, and F. Sattler. "Body core temperature sensing: Challenges and new sensor technologies". *Procedia Engineering*, 168, 89-92, 2016.
- [17] N. C. Jordan, J. H. Saleh, and D. J. Newman. "The extravehicular mobility unit: A review of environment, requirements, and design changes in the US spacesuit". *Acta Astronautica*, 59(12), 1135-1145, 2006.
- [18] P. Palacios, J. Cornejo, W. Castillo, M. V. Rivera, S. Tristan, J. Lezama and J.C. Chavez. "Telecommunications and Electronic Systems Analysis of T-EVA to Enhance the Body Temperature Monitoring during Extravehicular Activities on Mars Analog". In 2021 IEEE URUCON (pp. 294-298). IEEE, November, 2021.
- [19] P. Palacios, W. Castillo, M. V. Rivera, and J. Cornejo. "Design of T-EVA: Wearable Temperature Monitoring System for Upper Limbs during Extravehicular Activities on Mars". In 2020 IEEE Engineering International Research Conference (EIRCON) (pp. 1-4). IEEE, October, 2020.
- [20] M. V. Rivera, J. Cornejo, K. Huallpayunca, A. B. Diaz, Z. N. Ortiz-Benique, A. D. Reina, ... and V. Tiellacuri. "Medicina humana espacial: Performance fisiológico y contramedidas para mejorar la salud del astronauta". *Revista de la Facultad de Medicina Humana*, 20(2), 303-314, 2020.
- [21] T. T. Romet and J. Frim. "Physiological responses to fire fighting activities". *European journal of applied physiology and occupational physiology*, 56, 633-638., 1987.
- [22] I. Holmer and D. Gavhed. "Classification of metabolic and respiratory demands in fire fighting activity with extreme workloads". *Applied ergonomics*, 38(1), 45-52, 2007.

- [23] C. C. Roossien, R. Heus, M. F. Reneman, and G. J. Verkerke, "Monitoring core temperature of firefighters to validate a wearable non-invasive core thermometer in different types of protective clothing: Concurrent in-vivo validation". *Applied ergonomics*, 83, 103001, 2020.
- [24] V. Sandulescu and R. Dobrescu. "Wearable system for stress monitoring of firefighters in special missions". In 2015 E-Health and Bioengineering Conference (EHB) (pp. 1-4). IEEE. November, 2015.
- [25] D. L. Smith, S. J. Petruzzello, J. M. Kramer, and J. E. Misner, "The effects of different thermal environments on the physiological and psychological responses of firefighters to a training drill". *Ergonomics*, 40(4), 500-510, 1997.
- [26] R. E. Smith, R. W. Schutz, F. L. Smoll, and J. T. Ptacek. "Development and validation of a multidimensional measure of sport-specific psychological skills: The Athletic Coping Skills Inventory-28". *Journal of sport and exercise psychology*, 17(4), 379-398, 1995.
- [27] A. L. Bennett, J. Brown, A. Derchak, M. Di Marzo and S. T. Edwards. "Health and safety guidelines for firefighter training". Institute/Center for Firefighter Safety Research and Development, University of Maryland, 2006.
- [28] B. Belobrajdic, K. Melone, and A. Diaz-Artiles. "Planetary extravehicular activity (EVA) risk mitigation strategies for long-duration space missions". *NPJ Microgravity*, 7(1), 16, 2021.
- [29] J. P. Stroming and D. J. Newman. "Critical review of thermal management technologies for portable life support systems". 49th International Conference on Environmental Systems. July, 2019. Tdl.org. [Online]. Available: <https://ttu-ir.tdl.org/bitstream/handle/2346/84589/ICES-2019-338.pdf>. [Accessed: 20-Apr-2023].
- [30] M. G. Izenson, S. D. Phillips, A. B. Chepko, G. W. Daines, G. Quinn, and J. Steele. "Development of Lithium Chloride Absorber Radiator for Flight Demonstration. 47th International Conference on Environmental Systems". July, 2017. Tdl.org. [Online]. Available: [https://ttu-ir.tdl.org/bitstream/handle/2346/73073/ICES\\_2017\\_298.pdf?sequence=1](https://ttu-ir.tdl.org/bitstream/handle/2346/73073/ICES_2017_298.pdf?sequence=1). [Accessed: 30-Apr-2023].
- [31] N. C. Jordan, J. H. Saleh, and D. J. Newman. "The extravehicular mobility unit: A review of environment, requirements, and design changes in the US spacesuit. *Acta Astronautica*, 59(12), 1135-1145, 2006.
- [32] C. Chullen, I. Pena, K. Ganesan and H. Chen. "Advanced Technology Infusion into Spacesuit Systems". In ASCEND 2022 (p. 4351), 2022.
- [33] L. Kluis, N. Keller, N. Iyengar, H. Bai, R. Shepherd, and A. Diaz-Artiles. "An overview of the smartsuit architecture". 50th International Conference on Environmental Systems, July, 2021.
- [34] Center for Devices and Radiological Health, "3D printing of Medical Devices," U.S. Food and Drug Administration, 2022. [Online]. Available: <https://www.fda.gov/medical-devices/products-and-medical-procedures/3d-printing-medical-devices>. [Accessed: 01-Nov-2022].
- [35] J. Comejo, J. A. Comejo-Aguilar, M. Vargas, C. G. Helguero, R. Milanezi de Andrade, S. Torres-Montoya, ... and T. Russomano. "Anatomical Engineering and 3D printing for surgery and medical devices: International review and future exponential innovations". *BioMed research international*, 2022.
- [36] J. Charles and J. Railsback. "Human subsystem working group human planning guidelines and constraints," Nasa.gov, 2001. [Online]. Available: [https://history.nasa.gov/DPT/Human%20Exploration/Human%20Subsystems-Planning%20Guide%20JSC%20DPT%20Sept\\_01.pdf](https://history.nasa.gov/DPT/Human%20Exploration/Human%20Subsystems-Planning%20Guide%20JSC%20DPT%20Sept_01.pdf). [Accessed: 01-Nov-2022].
- [37] Maxim Integrated Products, Inc. "DS18B20 Programmable Resolution 1-Wire Digital Thermometer," Analog.com. [Online]. Available: <https://www.analog.com/media/en/technical-documentation/data-sheets/ds18b20.pdf>. [Accessed: 30-Apr-2023].
- [38] ESPRESSIF. "Datasheet ESP32 Series". 2023. Espressif.com. [Online]. Available: <https://www.espressif.com/sites/default/files/documentation/esp32-wroom-32>. [Accessed: 26-Oct-2022].
- [39] USB Enabling Connections TM. "USB type-C® cable and connector specification release 2.2". November, 2022. Usb.org. [Online]. Available: <https://www.usb.org/document-library/usb-type-cr-cable-and-connector-specification-release-22>. [Accessed: 26-Oct-2022].
- [40] Advanced Monolithic Systems. "AMS1117". Advanced-monolithic.com. [Online]. Available: <http://www.advanced-monolithic.com/pdf/ds1117.pdf>. [Accessed: 26-Oct-2022].
- [41] ILITEK. "a-Si TFT LCD Single Chip Driver 240RGBx320 Resolution and 262K color ILI9341 Specification" Adafruit.com. [Online]. Available: <https://cdn-shop.adafruit.com/datasheets/ILI9341.pdf>. [Accessed: 27-Oct-2022].
- [42] ESPRESSIF. "ESP-Now overview", 2023. Espressif.com. [Online]. Available in: <https://www.espressif.com/en/products/software/esp-now/overview>. [Accessed: 31-oct-2022]

# Artificial Intelligence-based Volleyball Target Detection and Behavior Recognition Method

Jieli Huang, Wenjun Zou\*

Physical education institute, Nanchang Jiaotong University, Nanchang 330100, China

**Abstract**—Volleyball has limitations in relying on judges' subjective judgments alone to call penalties for infractions in the court. While video detail enhancement technology is extremely useful for target tracking and extraction in sports video, the current research on video detail enhancement technology does not pay much attention to the development of ball game violation tracking and recognition. Therefore, the study uses the fusion algorithm of wavelet exchange method and three-frame difference method and background subtraction method to detect and extract the motion targets, and uses the improved CamShift tracking algorithm and HMM to track and identify the tracking targets for the violation actions. Comprehensively, the study constructs a tracking recognition model for volleyball violation based on video enhancement technology to achieve accurate penalty in intense rivalry games. Through experimental analysis and comparison, the tracking F-measure value of the model constructed by the study is 0.89, which can achieve a good tracking effect, the recognition accuracy is 99.76%, and the average error is 0.003, which can effectively realize the tracking recognition of players' illegal actions during volleyball, and objectively make court penalties to guarantee the fairness and justice of the game.

**Keywords**—Volleyball; video detail enhancement; hmm; CamShift tracking; detection; recognition

## I. INTRODUCTION

In competitive competitions, the subjective judgment of the referee may be influenced by different perspectives and observation results, leading to doubts about the accuracy of the judgment. Video detail enhancement technology has become an important research direction for achieving accurate punishment of illegal actions in volleyball matches. However, despite the rapid development of computer and network technology, significant progress has been made in related research fields; the application of video detail enhancement technology in sports video still faces some challenges [1]. Due to the uncertainty of lighting conditions at the competition site, limitations of collection equipment, and the presence of noise interference, the quality and clarity of volleyball game video images are often low. This poses a challenge to the effectiveness of video detail enhancement algorithms. In addition, as the competition progresses, moving targets may experience problems such as rapid movement, deformation, and occlusion, which also increases the difficulty of tracking and recognizing moving targets. In addition, due to the individual differences of different athletes and the complexity of game rules, further research is needed on the identification methods for illegal actions [2]. With the current computer and networking skills, video detail enhancement has matured. In

the process of image and video acquisition, due to the lighting environment, acquisition equipment, noise interference and other factors, it will lead to the degradation of image and video visual effects, so detail enhancement algorithms need to be used to highlight some of the detailed information in the image and video for subsequent processing [3]. There are two main types of video detail enhancement algorithms, namely, detail enhancement based on multiple input images and detail enhancement based on single input images [4]. While video detail enhancement techniques are extremely useful in the tracking and extraction of targets in motion video, not much research attention has been paid to video detail enhancement techniques [5]. The study uses video detail enhancement techniques to process video images in volleyball and to achieve detection and tracking of motion targets, and to identify the offending actions accordingly. The images are first preprocessed using the wavelet variation algorithm RGB. After pre-processing, the motion targets are detected and extracted using a fusion algorithm with three-frame difference method and background subtraction method. After the motion targets are detected and extracted, the motion targets are tracked and identified using CamShift tracking algorithm and Hidden Kolff (HMM), which is found to be lost when the target tracking may be obscured. To address this, the study adjusts the position of the CamShift algorithm's finding glass by introducing a Kalman filter (Kalman) to predict the motion parameters of the target. As a result, the study constructs a volleyball motion violation tracking recognition model based on video detail enhancement, which realizes the intelligence of game penalty and effectively solves the penalty problem caused by viewing angle and other reasons during the game. The importance of research lies in improving the accuracy and fairness of competition judgments, optimizing the effectiveness of video image processing and object tracking, and providing useful references for penalty issues in other sports. The innovation lies in the refinement of the application of video detail enhancement technology. Through the combination of various algorithms and models, a volleyball movement violation tracking and recognition model based on video detail enhancement has been constructed. This study has important practical significance in improving the accuracy of competition judgments. This study is mainly divided into five sections. The second section summarizes the research on video tracking, motion recognition, and other technologies both domestically and internationally. The third section is to construct a proposed volleyball foul action tracking and recognition model, analyzing image data processing, image detection, and target tracking. The fourth section is to analyze the performance of the constructed model and verify the

superiority of the proposed model. The fifth section discusses the results and analyzes deeper conclusions. The final section summarizes the results and proposes the shortcomings of the research and future research directions.

## II. RELATED WORKS

In the process of image and video acquisition, under the influence of lighting environment, acquisition equipment, noise interference and other factors, it will lead to the degradation of image and video visual effect, and need to use detail enhancement algorithm to highlight some information in the image and video for easy discrimination and processing. For this, many scholars conducted research to improve the image visual effect. Xue et al. [6] found that most video enhancement algorithms depend on optical flow to enroll frames in video sequences, but flow estimation is difficult, so they proposed a TOFlow to achieve enhancement of image data, and found excellent optimization by comparing three functional tasks, frame interpolation, video denoising/denoising, and video super-resolution, with conventional optical flow for standard benchmark tests. Guan et al. [7] proposed a MFQE method for the lower house video and designed a new MF-CNN to improve the quality of compressed video by addressing the problem that existing methods to improve the quality of compressed images/videos mainly focus on improving the quality of individual frames without considering the similarity between consecutive frames, which effectively improves the effectiveness and generalization in terms of the latest image quality of highly compliant videos. Wang et al. [8] found that the key challenge of video SR is to effectively exploit the correlated asphyxia among consecutive frames, and that available deep learning-based methods typically estimate the optical flow among LR frames to provide temporal dependence, proposing an end-to-end video SR network to super-resolve optical flow and image. Zheng et al. [9] constructed an unsaturated magnetic excitation-based MFL measurement device by converting MFL information to image representation through the Zaitong pseudo-color imaging protocol, and the maximum modulus method was used for the point location of wire breakage to extract color moments, statistical texture features, and spectral texture features from the image. Tang et al. [10] proposed a new ship detection model that can be called FLNet by combining image processing methods and deep learning target detection methods in order to solve the problem that there is a large amount of background information and noise information similar to the ship in the image due to the mechanism of imaging by SAR, which badly affects the ship detection model performance.

In intelligent video analysis systems, motion target tracking is widely used in intelligent surveillance, human-computer interaction, and autonomous driving. In order to be able to track and recognize with high accuracy despite the challenges such as environmental changes, occlusion deformation and scale changes of the tracking target, the research on motion tracking recognition has been increasing. Kim et al. [11] tracked the excavator by using a pre-trained detector to locate the excavator and correlate the detection results, tracked the excavator by the tracking learning clean toilet algorithm, and finally used a hybrid deep

learning algorithm to model the visual features of the excavator and the operation cycle of the sequential patterns were modeled and trained to propose a vision-based framework for animal recognition. Jaouedi et al. [12] proposed a hybrid deep learning model-based approach for human action recognition using gated recurrent neural networks to classify sequence data and videos, and upper and lower feature data extraction using evaluation algorithm. Angeliniet al. [13] proposed a hybrid deep learning model-based approach for HAR due to the gap between the deep learning data requirements and the functionality provided by the framework that needs to provide the application in terms of data recording devices, a new 2D pose based pose level HAR approach (ActionXPose) was proposed by reducing the gap using the human pose provided by OpenPos. Zhang et al. [14] proposed a new pose level HAR approach (ActionXPose) in order to deal with the gaps with different temporal context information for long duration time series features and enhance spatio-temporal attention, the human action recognition problem was solved by using Conv-LSTM and FC-LSTM, and a STDAN was designed. Ge et al. [15] effectively represent the spatial static and temporal dynamic information of videos, using GoogleNet to extract the features in the video frames, processed by a spatial transformer network then modeled the sequential information of the features by convolutional LSTM, and proposed a new attention mechanism based convolutional LSTM action recognition algorithm.

In summary, the CNN model is used as the mainstream direction in speech recognition technology. Although this model can improve the performance of speech recognition, the subsequent structure construction is complicated and is not conducive to improving the operation efficiency of the model. Therefore, the research starts from neural network and proposes OPGRU to simplify the structure of speech recognition model and improve the recognition accuracy and operation efficiency of the model.

## III. VOLLEYBALL FOUL ACTION TRACKING RECOGNITION MODEL CONSTRUCTION

### A. Image Data Pre-Processing and Target Detection

Common penalty errors and misjudgments have been the trigger for conflicts on the court. The requirement of action specification in volleyball is very high, in order to ensure the reduction of errors and wrong calls in volleyball and the accurate detection of fouls committed by players. The ball game video is processed and analyzed using video detail enhancement technology [16]. Before tracking and recognition of ball game violation actions, pre-processing of sports video image data is required to improve the accuracy and recognition precision of tracking afterwards. Firstly, the image is converted to grayscale image by RGB and the image is weighted, and the processing formula is shown in the following Eq. (1).

$$I_i(x, y) = \alpha r(x, y) + \beta g(x, y) + \gamma b(x, y) \quad (1)$$

The values  $\alpha$ ,  $\beta$ , and  $\gamma$  in Eq. (1) are the values for weighting the action images of volleyball players. When the



athlete action image is an action image captured in natural light is, the image weight is set to 1. When the action image is captured in a single light, the image weight is set to 0 to eliminate the shadow of the volleyball player's body, and the wavelet transform algorithm is used to reduce the noise of the data signal after the image is grayscale processed and weighted, and the wavelet transform formula is shown in Eq. (2) below.

$$WT(a, \tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t) \cdot \varphi\left(\frac{t-\tau}{a}\right) dt \quad (2)$$

In Eq. (2),  $a$  denotes the variable scale and  $\tau$  denotes the variable translation. After multiple wavelet transforms, the obtained signal components are shown in Eq. (3) below.

$$S = A_n + D_1 + D_2 + \dots + D_n \quad (3)$$

In Eq. (3),  $S$  denotes the original signal,  $D_n$  denotes the noisy signal obtained after  $n$  wavelet transforms, and  $A_n$  denotes the effective signal obtained after  $n$  wavelet transforms. After verification, it is found that the 3-layer wavelet transform has the best denoising effect. According to the above, the denoising and grayscale processing of volleyball sports video action data is realized. After processing the image need to detect the motion target, motion target detection is to take the target color, shape, position and size information in each frame of the video stream, and the video sequence is essentially three-dimensional data containing a time dimension, the study uses the three-frame difference method to extract the motion target, the three-frame difference method is illustrated in Fig. 1.

In Eq. (8),  $\alpha$  is the number, which is fed into the section as needed,  $w$  is the degree, and  $h$  is the figure. Eq. (9) is obtained by "summing" the background subtraction method with the motion target information obtained from the three-frame difference method.

$$D(x, y) = DI(x, y) \otimes DB(x, y) \quad (9)$$

After obtaining the target information, the background adaptive update is performed, and the update expression is shown in Eq. (10).

$$B(x, y) = \begin{cases} B(x, y) & D(x, y) = 0 \\ uB(x, y) + (1-u)I_k(x, y) & D(x, y) \neq 0 \end{cases} \quad (10)$$

The best value of  $u$  in Eq. (10) is 0.997, which is obtained after the study. The flow chart of the fusion algorithm of the three-frame difference method and the background subtraction method is shown in Fig. 2.

Firstly, the three adjacent frame degree values  $I_{k-2}$ ,  $I_{k-1}$  and  $I_k$  are collected for the operation of the neighboring two

difference absolute, and the difference map formula is obtained as shown in Eq. (4).

$$\begin{cases} D_1(x, y) = |I_{k-1} - I_{k-2}| \\ D_2(x, y) = |I_k - I_{k-1}| \end{cases} \quad (4)$$

Binarizing the two neighbor-dual differences, the expression is shown in Eq. (5).

$$\begin{cases} T_1 = d_1 + \beta\delta_1 \\ T_2 = d_2 + \beta\delta_2 \\ D_1(x, y) = \begin{cases} 255 & d_1 \geq T_1 \\ 0 & d_1 < T_1 \end{cases} \\ D_2(x, y) = \begin{cases} 255 & d_2 \geq T_2 \\ 0 & d_2 < T_2 \end{cases} \end{cases} \quad (5)$$

In Eq. (5),  $d$  is the mean value of the difference map,  $\delta$  is the standard deviation of the difference map, and  $T$  represents the threshold value. The two binarized images are subjected to or operation to obtain the motion target information, as shown in Eq. (6) below.

$$DI(x, y) = D_1(x, y) \oplus D_2(x, y) \quad (6)$$

The above equation is used to process the image, but in the process of research, it is found that there are still some limitations, and the motion target in the extraction will produce a hole phenomenon, for this problem, the study combines the background subtraction method and the three-frame difference method. The Eq. (7) is obtained from the background subtraction method.

$$DB(x, y) = |I_k(x, y) - B(x, y)| \quad (7)$$

In Eq. (7)  $I_k(x, y)$  is the current degree value,  $B(x, y)$  is the back pixel gray value.  $d(x, y)$  is the difference absolute value image  $DB(x, y)$  value image  $DB(x, y)$  pixel point of  $\bar{d}$  and standard  $\delta$ , with and standard deviation set to a value of  $T$ , into the binarization, to obtain the expression (8).

$$\begin{cases} \bar{d} = \frac{\sum_{x=0}^{w-1} \sum_{y=0}^{h-1} d(x, y)}{wh} \\ \delta = \sqrt{\frac{\sum_{x=0}^{w-1} \sum_{y=0}^{h-1} [d(x, y) - \bar{d}]^2}{wh}} \\ T = \bar{d} + \alpha\delta \\ DB(x, y) = \begin{cases} 225 & d \geq T \\ 0 & d < T \end{cases} \end{cases} \quad (8)$$

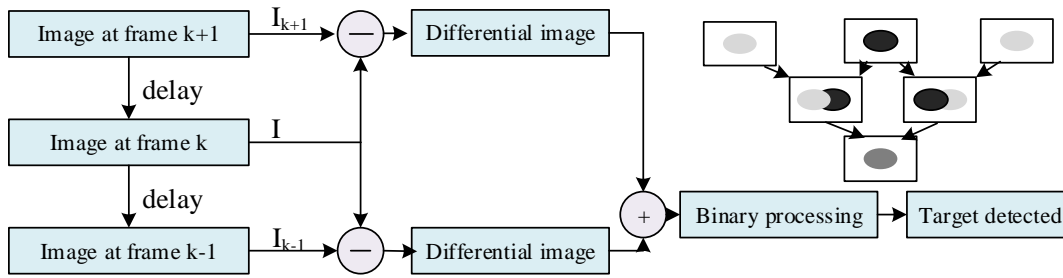


Fig. 1. Diagram of three frame difference method.

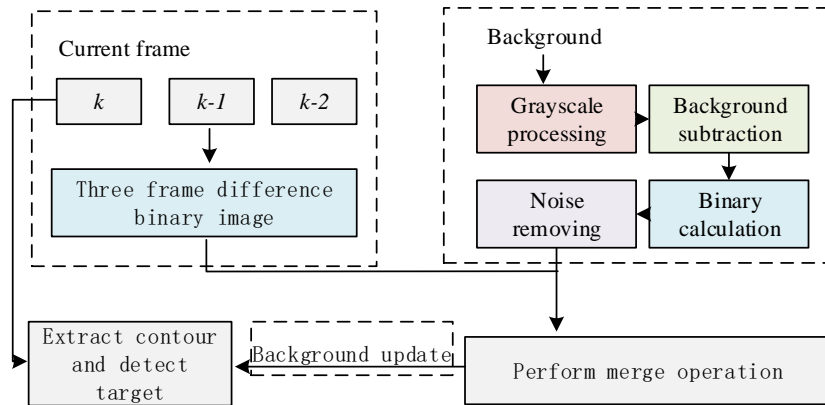


Fig. 2. Flow chart of fusion algorithm for target detection.

Combining the above, the study implements the pre-processing of volleyball sports action video images, and detects and extracts the image targets using an algorithm fused with the three-frame difference method and the background elimination method.

### B. Target Tracking based on Improved CamShift Algorithm

After the target is detected and extracted, it needs to be tracked so that the subsequent violation movements during volleyball sports can be identified in real time. The CamShift is a successive self-adaptive Meanshift algorithm. The Meanshift algorithm that finds and iterates over the pixels of a single image for optimal results, the CamShift mainly iterates processing of video sequences, where each frame in the video is called using the Meanshift algorithm. The Meanshift algorithm belongs to the kernel density estimation method, which describes the model of the target and the candidate model by the probability of the pixel feature values in the region and in the candidate region. Given a sample of points in the d-dimensional space  $R^d$ , the Meanshift vector of points has the basic form shown in Eq. (11).

$$M_h(x) = \frac{1}{k} \sum_{x_i \in S_h} (x_i - x) \quad (11)$$

In Eq. (11),  $x_i$  is the sample points in the interval,  $k$  denotes the sample falling into the  $S_h$  region, and  $S_h$  is the high-dimensional spherical region of radius  $h$ , which is the set of  $y$  points satisfying the relationship in Eq. (12) below.

$$S_h(x) = \{y : (y-x)^T (y-x) \leq h^2\} \quad (12)$$

The study extended the basic Meanshift form by introducing a kernel function in order to take into account the effect of the distance of each pixel point during the calculation, as shown in Eq. (13) below.

$$M_h(x) = \frac{\sum_{i=1}^n G\left(\frac{x_i - x}{h}\right) \omega(x_i) (x_i - x)}{\sum_{i=1}^n G\left(\frac{x_i - x}{h}\right) \omega(x_i)} \quad (13)$$

In Eq. (13)  $G(x)$  is a unit kernel function and  $\omega$  is the weight assigned to the sampled points. Using Eq. (13) for iteration, the following Eq. (14) is obtained.

$$m_h(x) = M_h(x) + x \quad (14)$$

After calculating the value of  $m_h(x)$ , assign it to  $x$ , and then calculate  $M_h(x)$  again. If the absolute value of  $M_h(x)$  is less than the tolerance error, end the cycle to get the final target position. If not, continue the calculation. The meanshift tracking algorithm is shown in Fig. 3.

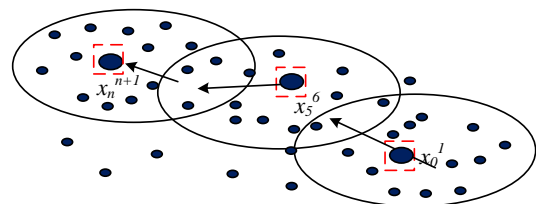


Fig. 3. Schematic diagram of MeanShift tracking algorithm.

The CamShift utilizes the modalities of the probability distribution image detected by the Meanshift algorithm in continuous video detection by introducing a feedback loop in which the previous detection result is used as input to the next detection process and restricting the search area to the surroundings of the latest known target location. After the color tracking probability model is established, the video image is transformed with a color probability distribution map. The model initializes a search window (SW) in the first frame of the image, and adjusts the window size and position according to the tracking target in the way shown in Eq. (15).

$$\begin{cases} \hat{p}_k(W) = \frac{1}{|W|} \sum_{j \in W} p_j \\ \hat{p}_k(W) - p_k \approx \frac{f'(p_k)}{f(p_k)} \end{cases} \quad (15)$$

In Eq. (15)  $W$  is the SW of size  $s$  in the color probability distribution map,  $p_k$  is the initialized centroid of the SW,  $f(p)$  is the Meanshift climbing gradient equation, and the new centroid  $\hat{p}_k$  is found by dynamic iteration, and this is cycled until convergence to achieve adaptive change of the window. During the study, it is found that the target tracking may be obscured during the target tracking resulting in target loss. To address this, the study adjusts the position of the SW of the CamShift by introducing the Kalman filter (Kalman) to predict the motion parameters of the target to compensate for the temporary target loss due to occlusion. In two stages, prediction and correction, the components of the predicted state vector are used to set the center position of the SW of the CamShift, and the center of gravity position output by the CamShift is used as the measurement value to correct the predicted state vector and achieve the optimization improvement of the CamShift.

### C. HMM-based Foul Play Analysis Identification

Judgment of foul actions and near-foul actions in volleyball by the eyes of the referee alone usually leads to errors. The study has already used a fusion algorithm combining the three-frame difference method and the background elimination method to detect and extract the motion targets in the previous paper, and the improved CamShift has been used to achieve the tracking of motion targets in ball game sports videos. The HMM recognition process is shown in Fig. 4.

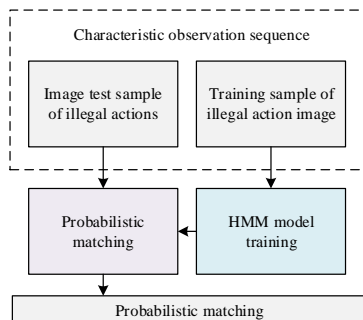


Fig. 4. HMM violation action identification process.

The data set of foul actions of volleyball players is divided into different types, firstly, each violation is modeled in a targeted way, and the study sets the amount of data for each violation to 120 and decomposes the action into  $n$  meta-actions. These meta-actions during the motion are temporal in nature, so each game violation is considered as a sequence of observations of length  $n$ . Training and learning are performed for this sequence of observations to find the best HMM parameters for each action model [17]. After finding the best parameters, the extracted observation sequence data is used as the input data of the HMM, and the probability of the best state sequence occurrence of the action in the current video under each action model is obtained using the Viterbi algorithm. The action which corresponds to the model with the largest probability of output is the identification outcome of the current observed sequence. HMM is expressed as the following Eq. (16).

$$\lambda = (A, B, \pi) \quad (16)$$

In Eq. (16),  $A$  is the state probability distribution,  $B$  is the observation probability distribution, and  $\pi$  is the initial probability distribution. To find the appropriate HMM parameters, the study uses the Baum-Welch algorithm to train each parameter of the HMM so that the probability of the observed sequence in the model is maximized. The state sequence data is considered as unobservable hidden data  $I$  as shown in Eq. (17) below.

$$P(O|\lambda) = \sum_I P(O|I, \lambda) P(I|\lambda) \quad (17)$$

In Eq. (17),  $O$  is the observed sequence data. The maximum expectation algorithm (EM) is used to implement the HMM algorithm for parameter learning.  $Q$  function, as follows in Eq. (18).

$$Q(\lambda, \bar{\lambda}) = \sum_I \log P(O, I|\lambda) P(O, I|\bar{\lambda}) \quad (18)$$

In Eq. (18),  $\bar{\lambda}$  denotes the current estimate of the model parameters and  $\lambda$  is the maximized model parameters. After obtaining the value of the  $Q$  function, the parameters of the HMM were obtained by maximizing the  $Q$  function, combined with the Lagrange multiplier method. In volleyball violation recognition, after the training of the violation model is completed, the study uses the Viterbi algorithm to find the optimal solution of the HMM. For a given HMM model and observed sequence data, the optimal path  $I^* = (i_1^*, i_2^*, \dots, i_T^*)$  is solved, and  $T$  denotes the length of sequence  $I$ . Through the above operation, the analysis of the recognition of ball game violation actions is completed. The principle of the Viterbi algorithm is shown in Fig. 5.

Comprehensive above, the research uses video detail enhancement technology to detect and track the target of the game video image, and finally uses the action model for recognition, and constructs a volleyball foul action tracking recognition model based on video detail enhancement technology, which effectively makes accurate judgment on each violation action remembered in volleyball.

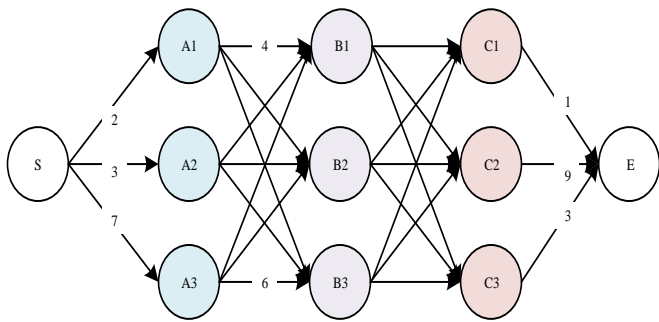


Fig. 5. Principle of viterbi algorithm.

#### IV. TRACKING RECOGNITION MODEL PERFORMANCE ANALYSIS

The study conducted performance analysis on the constructed model, using the Volleyball Dataset for training and analysis. The dataset consists of 55 videos, with 4830 keyframes annotated with athlete positions, as well as their individual and group actions. The study divided 4830 keyframes into training and testing sets, with a ratio of 8:2 between the training and testing sets. The study was conducted to improve the accuracy and recognition precision of tracking. The wavelet transform noise reduction process was performed on the image, and to verify the noise reduction process effect, the study used the same set of violating actions. The noise reduction effect of the image of the x-axis acceleration of the violation action is compared as shown in Fig. 6.

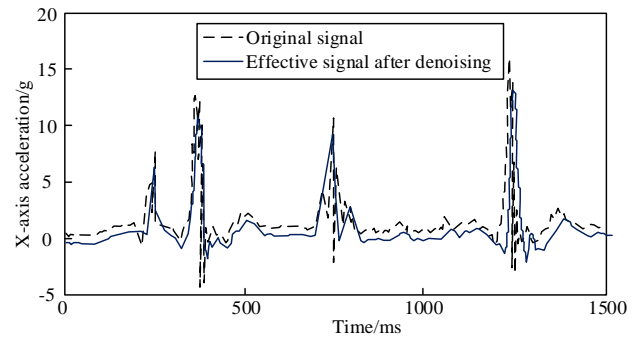


Fig. 6. Effect before and after wavelet change noise reduction.

As can be seen from Fig. 6, before the use of wavelet denoising, there is a lot of redundant data and noise in the image of the offending action, which is difficult to extract and identify subsequently, after the wavelet transform noise reduction eliminates the redundant data removal in the x-axis acceleration of the foul action, and maintains the original curve direction while noise reduction, which makes the waveform graph clearer and improves the accuracy of tracking and identification of subsequent video images. To verify the tracking effect of the tracking algorithm, the research improved CamShift tracking algorithm is compared longitudinally with the currently popular and superior performance tracking algorithms: CT, TLD, IVT, and LIPAG to verify the tracking of a subset of target deformation, a subset of illumination changes, and a subset of background interference, a subset of the three interference cases using the algorithm. The distance accuracy curves (precision plot) for comparing different cases are shown in Fig. 7.

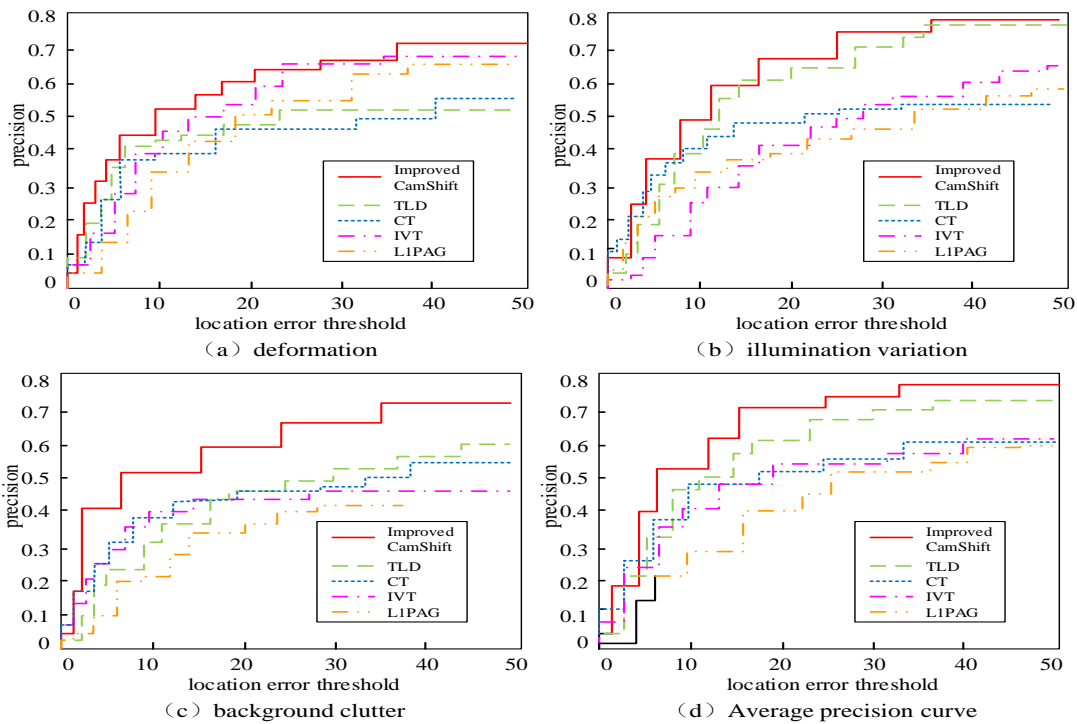


Fig. 7. Comparison of Accuracy Curves.

In Fig. 7, the tracking distance precision of each algorithm is affected by different effects under different interference scenarios, and increases with the increase of the positioning error threshold, among which the improved CamShift tracking algorithm is the least affected by interference, and the average tracking precision is 0.79 under each interference scenario at a positioning error threshold of 50. The TLD tracking algorithm is at a positioning error threshold of 50. The average tracking precision of the TLD tracking algorithm is 0.73 at a positioning error threshold of 50, which is 8.2% less than the precision of the improved structure. The average tracking precision of the CT tracking algorithm is 0.59 at a positioning error threshold of 50, which is 33.9% less than the precision of the improved structure. The average tracking precision of the IVT tracking algorithm is 0.61 at a positioning error threshold of 50, which is 29.5% less than the precision of the improved structure. The average tracking precision of the LIPAG tracking algorithm is 0.59 at a positioning error threshold of 50, which is 29.5% less than the precision of the improved structure; and the average tracking precision of the LIPAG tracking algorithm is 0.61 at a positioning error threshold of 50. The average tracking precision of the LIPAG tracking algorithm is 5.7 when the localization error threshold is 50, which is 38.6% less than the algorithm precision. The above figure shows that the proposed algorithm has better tracking effect in complex scenarios such as target deformation, light change, and background disturbance. To further verify the performance of the algorithms, the study introduces recall (Re), precision (Pr), and comprehensive performance (F-measure) to compare the performance of the five tracking algorithms under three different scenarios of multimodal background, light change, and bad weather with quantitative metrics, as shown in Table I.

TABLE I. COMPARISON OF AVERAGE PERFORMANCE OF TRACKING ALGORITHMS

Algorithm	Scene	Re	Pr	F-measure
Improved CamShift	Highway	0.90	0.88	0.89
	Fountain	0.88	0.87	0.87
	Wet Snow	0.92	0.91	0.91
TLD	Highway	0.82	0.81	0.81
	Fountain	0.84	0.80	0.82
	Wet Snow	0.76	0.79	0.78
CT	Highway	0.82	0.83	0.82
	Fountain	0.79	0.80	0.79
	Wet Snow	0.80	0.81	0.80
IVT	Highway	0.73	0.75	0.74
	Fountain	0.69	0.70	0.69
	Wet Snow	0.77	0.79	0.78
LIPAG	Highway	0.74	0.72	0.73
	Fountain	0.71	0.74	0.72
	Wet Snow	0.68	0.69	0.68

Table I shows that the improved CamShift tracking algorithm using particle filtering significantly improves the ability of the improved algorithm to handle complex backgrounds including light changes and multimodal backgrounds, and the performance index value of this algorithm is the highest in all scenes. The improved CamShift tracking algorithm has an average Re value of 0.90, an average Pr value of 0.89, and an F-measure value of 0.89 for the three scenes. Slightly higher than the TLD tracking

algorithm, which has an average Re value of 0.81, an average Pr value of 0.80, and an F-measure value of 0.80. The CT tracking algorithm has an average Re value of 0.81. The average Re value of the CT tracking algorithm is 0.81, the average Pr value is 0.81, and the F-measure value is 0.81; the other two algorithms have lower values of quantitative indicators. The comprehensive table above shows that the research improved CamShift tracking algorithm has high comprehensive performance and good robustness. To verify the recognition effect of the volleyball violation recognition model (model 1) constructed by the study, the recognition models constructed by convolutional neural network (CNN) (model 2) and support vector machine (SVM) (model 3) and BP neural network (model 4) were used for violation recognition using volleyball sports videos from video websites, and the recognition effect was compared for training and testing, and the results are shown in Fig. 8 is shown.

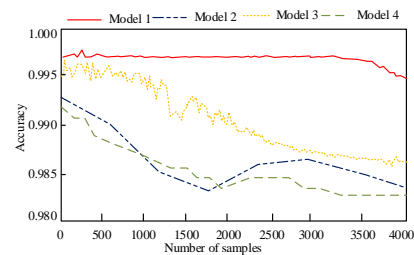


Fig. 8. Comparison of recognition effects of four recognition models.

In Fig. 8, the recognition rate of the offending actions in models 1, 2, 3, and 4 decreases as the sample size increases. When the sample reaches at 500, the precision of model 1 was 99.76%, the recognition precision of model 3 was 98.68% lower than that of model 1 by 0.08%, the recognition precision of model 2 was 99.21%, lower than that of model 1 by 0.55%, and the recognition precision of model 4 was 98.89%, lower than that of model 1 by 0.87%. When the sample was increased to 4000, the precision of the four model precision all decreased, but model 1 decreased the least, the recognition precision was 99.52%, model 3 recognition precision was 98.39%, 1.13% lower than model 1, model 2 recognition precision was 98.49%, 1.03% lower than model 1, model 4 recognition precision was 98.32%, 1.20% lower than model 1. The comprehensive content of Fig. 6 shows that model 1 has high stability and the highest recognition precision among the four recognition models. To further verify the recognition model performance, the recognition errors of the four models are recorded and compared, as shown in Fig. 9.

From Fig. 9, the highest error of recognition error curve of model 1 is 0.009, the lowest error is 0.001, and the average error is 0.003. It can be seen that the overall curve of this model is lower than the other three models, among which the highest recognition error of model 3 is 0.012, the highest recognition error of model 4 is 0.018, and the highest recognition error of model 2 is 0.016. The volleyball violation tracking recognition model constructed by the study can effectively track and identify the violations during the game, providing a strong guarantee for the fairness of the game.

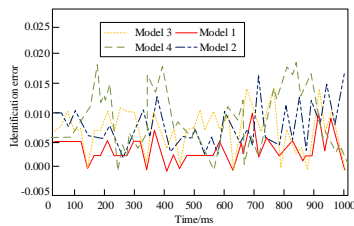


Fig. 9. Identification error of four models.

## V. DISCUSSION

Through the above experimental results, it was found that wavelet transform denoising of images can effectively eliminate redundant data and noise, maintain the original curve direction, and make the waveform clearer. This is crucial for the tracking and recognition of subsequent video images. Compared with popular tracking algorithms such as CT, TLD, IVT, and LIPAG, the improved CamShift tracking algorithm has higher tracking accuracy and robustness in subsets of illumination changes, background interference, and target deformation. Similarly, accuracy, recall, and comprehensive performance indicators also demonstrate its superiority. Compared with popular tracking algorithms such as CT, TLD, IVT, and LIPAG, the improved CamShift tracking algorithm has higher tracking accuracy and robustness in subsets of illumination changes, background interference, and target deformation. Similarly, accuracy, recall, and comprehensive performance indicators also demonstrate its superiority. The recognition error of Model 1 is the smallest, indicating that the model has high robustness and accuracy, and is suitable for use in actual competitions to accurately track and identify violations that occur during the competition. In summary, the improved CamShift tracking algorithm and volleyball violation recognition model constructed in the study have been compared and tested, showing high accuracy and stability. This will have a significant impact on the fairness of actual matches and open up a path for subsequent research, indicating that important positions should be given to data preprocessing and model optimization in such research.

## VI. CONCLUSION

In volleyball, referees are very prone to subjective judgment errors due to different observations from different angles during the game viewing process. For this reason, the study uses RGB to grayscale the image and wavelet variation algorithm to noise reduce the video image, and then uses the fusion algorithm of three-frame difference method and background subtraction method to detect and extract the motion target in the image. After the motion target is detected and extracted, the CamShift tracking algorithm is used to track the motion target and it is found that the target may be lost due to the occlusion of the tracking target during the target tracking. To address this, the study adjusts the position of the SW of the CamShift by introducing the motion parameters of the Kalman prediction target to achieve improved optimization. The set of images that have been extracted from the offending actions are input to the HMM recognition model to track and identify the offending actions that appear in the motion video. As a result, the study constructs a volleyball

motion violation tracking recognition model based on video detail enhancement. Through the analysis of experimental verification, the recognition accuracy of the research constructed tracking recognition model is 99.76%, and the average error is 0.003, which can effectively realize the tracking recognition of players' illegal actions during volleyball sports and realize the fairness of the penalty in the court game. At present, the model is only used in sports competitions and has not been put into more fields, which can be further explored in the future research.

## REFERENCE

- [1] Hesser B. The protection of minor athletes in sports investigation proceedings. *The International Sports Law Journal*, 2021, 21(1):62-73.
- [2] Bao W, Lai W S, Zhang X, Gao Z, Yang M H. Memc-net: Motion estimation and motion compensation driven neural network for video interpolation and enhancement. *IEEE transactions on pattern analysis and machine intelligence*, 2019, 43(3):933-948.
- [3] Liu R, Fan X, Zhu M, Hou M, Luo Z. Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light. *IEEE Transactions on Circuits and Systems for Video Technology*, 2020, 30(12):4861-4875.
- [4] Wang W, Chen Z, Yuan X, Wu X. Adaptive image enhancement method for correcting low-illumination images. *Information Sciences*, 2019, 496:25-41.
- [5] Hou R, Zhou D, Nie R, Liu D, Xiong L, Guo Y, Yu C. VIF-Net: an unsupervised framework for infrared and visible image fusion. *IEEE Transactions on Computational Imaging*, 2020, 6: 640-651.
- [6] Xue T, Chen B, Wu J, Wei D, Freeman W T. Video enhancement with task-oriented flow. *International Journal of Computer Vision*, 2019, 127(8): 1106-1125.
- [7] Guan Z, Xing Q, Xu M, Yang R, Liu T, Wang Z. MFQE 2.0: A new approach for multi-frame quality enhancement on compressed video. *IEEE transactions on pattern analysis and machine intelligence*, 2019, 43(3): 949-963.
- [8] Wang L, Guo Y, Liu L, et al. Deep video super-resolution using HR optical flow estimation. *IEEE Transactions on Image Processing*, 2020, 29: 4323-4336.
- [9] Zheng P, Zhang J. Quantitative nondestructive testing of wire rope based on pseudo-color image enhancement technology. *Nondestructive Testing and Evaluation*, 2019, 34(3): 221-242.
- [10] Tang G, Zhao H, Claramunt C, et al. FLNet: A Near-shore Ship Detection Method Based on Image Enhancement Technology. *Remote Sensing*, 2022, 14(19): 4857-4857.
- [11] Kim J, Chi S. Action recognition of earthmoving excavators based on sequential pattern analysis of visual features and operation cycles. *Automation in Construction*, 2019, 104: 255-264.
- [12] Jaouedi N, Boujnah N, Bouhlel M S. A new hybrid deep learning model for human action recognition. *Journal of King Saud University-Computer and Information Sciences*, 2020, 32(4): 447-453.
- [13] Angelini F, Fu Z, Long Y, Shao L, Naqvi S M. 2D pose-based real-time human action recognition with occlusion-handling. *IEEE Transactions on Multimedia*, 2019, 22(6): 1433-1446.
- [14] Zhang Z, Lv Z, Gan C, Zhu Q. Human action recognition using convolutional LSTM and fully-connected LSTM with different attentions. *Neurocomputing*, 2020, 410: 304-316.
- [15] Ge H, Yan Z, Yu W, et al. An attention mechanism based convolutional LSTM network for video action recognition. *Multimedia Tools and Applications*, 2019, 78(14): 20533-20556.
- [16] Liu N, Liu P. Goaling recognition based on intelligent analysis of real-time basketball image of Internet of Things. *Journal of supercomputing*, 2022, 78(1):123-143.
- [17] Wang Y, Song Q, Ma T, Yao N, Liu R, Wang B. Research on human gait phase recognition algorithm based on multi-source information fusion. *Electronics*, 2022, 12(1): 193-193.

# Deep Learning-based Multiple Bleeding Detection in Wireless Capsule Endoscopy

Prof. Ouiem Bchir, Ghaida Ali Alkhudhair, Lena Saleh Alotaibi,  
Noura Abdulhakeem Almhizea, Sara Mohammed Almuhanha, Shouq Fahad Alzeer  
Collage of Computer Science and Information, King Saud University, Riyadh, Kingdom of Saudi Arabia

**Abstract**—Wireless Capsule Endoscopy (WCE) is a diagnostic technology for gastrointestinal tract pathology detection. It has emerged as an alternative to conventional endoscopy which could be distressing to the patient. However, the diagnosis process requires to view and analyze hundreds of frames extracted from WCE video. This makes the diagnosis tedious. For this purpose, researches related to the automatic detection of signs of gastrointestinal diseases have been boosted. In this paper, we design a pattern recognition system for detecting Multiple Bleeding Spots (MBS) using WCE video. The proposed system relies on the Deep Learning approach to accurately recognize multiple bleeding spots in the gastrointestinal tract. Specifically, the You Only Look Once (YOLO) Deep Learning models are explored in this paper, namely, YOLOv3, YOLOv4, YOLOv5 and YOLOv7. The results of experiments showed that YOLOv7 is the most appropriate model for designing the proposed MBS detection system. Specifically, the proposed system achieved a mAP of 0.86, and an IoU of 0.8. Moreover, the results of the detection were enhanced by augmenting the training data to reach a mAP of 0.883.

**Keywords**—Wireless Capsule Endoscopy (WCE); Multiple Bleeding Spots (MBS); Gastrointestinal (GI) disease; deep learning; pattern recognition

## I. INTRODUCTION

The digestive system disorders have been a concern for physicians over years. In fact, millions of people around the world suffer from gastrointestinal (GI) diseases. Specifically, among more than 73 thousand participants in a worldwide study, 40% of them have functional gastrointestinal disorders. In addition, disorders such as digestive system cancer are considered fatal and a major cause of mortality according to 2020 United States statistics. Several pathogens can affect the gastrointestinal tract such as inflammations, infections, cancers, benign tumors, ulcers, and hemorrhoids. Some of these pathogens have similar symptoms. Specifically, cancer, benign tumors, ulcers, and hemorrhoids may yield Multiple Bleeding Spots (MBS) in the gastrointestinal tract. The latter symptom consists of a loss of blood in the GI tract because of ruptured vessels indicating the presence of an abnormality [1]. These MBS appear as small dark red spots or as small light spots next to the red dark ones. Fortunately, with the emergence of new diagnostic techniques, it is possible for physicians to detect GI abnormalities. Endoscopy is the most common diagnostic technique for GI tract. Nevertheless, it is inconvenient and painful for the patient. In order to alleviate this inconvenience, Wireless Capsule Endoscopy (WCE) developed in 2000, emerged as a new diagnostic technique.

The diagnosing process consists of the patient swallowing a capsule. The latter contains a camera to record the journey of the capsule internally to the GI tract. Then, the physician analyses the record to diagnose the patient by looking for abnormal spots. WCE generates an eight-hour video. In other words, 60,000 frames need to be visualized by the physician. However, due to the small size of the lesion region and the visual fatigue, the disease diagnosis may be missed at an early stage. In light of this, a diagnostic technology related to image processing and pattern recognition would help in the rapid and accurate detection of the disease. Nevertheless, due to the likeness of the MBS and other intestinal characteristics such as, bubbles, holes, or small food debris, etc. It is challenging to extract visual descriptors able to distinguish MBS pattern from the other ones. It is even more arduous due to the background clutter. In fact, MBS can occur in all parts of the GI tract exhibiting large variety of background in terms of color, and texture. One way to tackle this problem is through the use of Deep Learning (DL) models which learn automatically suitable features.

In this paper, we develop a multiple bleeding spot detection system for Wireless Capsule Endoscopy (WCE) videos. More specifically, we design a pattern recognition system based on deep learning models that are able to detect the bleeding spots through the GI tract. In particular, deep learning models adopted for pattern recognition were utilized. These models are designed to localize and categorize the object of interest. For this purpose, we employ the You Only Look Once (YOLO) deep learning approach [2]. In this regard, we propose to compare different versions of YOLO. These are YOLOv3 [3], YOLOv4 [4], YOLOv5 [5], and YOLOv7 [6].

## II. RELATED WORKS

Recent researches have proposed aided-diagnosis systems for bleeding anomalies within the intestinal tract using WCE images. They can be categorized into classification-based approaches, and detection-based approaches. The former approaches classify the whole WCE frame as including bleeding spots or not including bleeding spots. Whereas, the detection approaches not only classify the frame but also localize the bleeding spots within the frame. Moreover, each of these two categories bifurcates into conventional and deep learning approaches according to the machine learning paradigm that have been adopted. More specifically, conventional approaches use “engineered” features (also referred to as hand crafted features). Alternatively, deep

learning approaches automatically extract the feature while training the deep learning model.

#### A. WCE Frame Classification System

1) *Conventional approaches:* The work in [7] propose to classify WCE frames into “Bleeding” and “No Bleeding”. For this purpose, it extracts a hand-crafted feature, namely, the color moment feature from WCE frames. Then, it is conveyed to a Support Vector Machine (SVM) [8] classifier. The choice of the visual feature to be adopted has been made through empirical experimentation. In fact, MPEG-7 visual descriptors, “color moment”, “Discrete Wavelet Transform”, “Edge Histogram Descriptor”, “Gabor”, and a combination of “Discrete Wavelet Transform” and “color moment”. Similarly, the proposed system in [9] extracts hand crafted features. More specifically, MPEG-7 features [10] are considered. These are the “color moments”, the “color histogram”, the “local color moments”, the “Gabor filter”, the “Discrete Wavelet Transform” (DWT) and the “Local Binary Pattern” (LBP) features [10]. The extracted features are then conveyed to a machine learning approach to categorize the frames as “Bleeding” or “No Bleeding”. This is performed by clustering each of the training “Bleeding” frames, and the training “No Bleeding” frames into similar groups using Fuzzy C-Means (FCM) [11]. As such, in the testing phase, the unknown frame is compared to the obtained cluster centroids from the training phase. It is then assigned to class of the closest centroid.

2) *Deep learning approaches:* The authors in [12] use a well-known CNN model that won of the ImageNet Large Scale Vision Recognition Competition (ILSVRC). Specifically, it exploits LeNet-5 [13] architecture. Alternatively, the work in [14] uses deep learning CNN models for feature extraction. In particular, VGG-19 [15], ResNet50 [16], and InceptionV3 [17] are adopted. Similarly, these are well known CNN models which won the ILSVRC competition. Nevertheless, inceptionV3 is an evolved version of InceptionV1 used in GoogleNet displays the architecture of inceptionV3. The obtained features from the three considered models are concatenated. Then, a feature selection is performed to select the most distinctive features. The selected features are conveyed to SVM classifier [8] to categorize the frames as “Bleeding” or “No Bleeding”. The study in [18] proposed a system to diagnose the abnormalities in the GI. This study proposed a model which utilizes MobileNet [19]. The latter is a lightweight deep learning model. Specifically, it uses the independent convolutions for each depth dimension, then employs  $1 \times 1$  pointwise convolution to recover the depth. The output of MobileNet [19] is fed to a custom built convolutional neural network model. It is constituted of 64 filters with a kernel size of  $3 \times 3$ . The resulting feature map is passed to a three fully connected layers for classification purpose. In [20] authors proposed to classify WCE frames as “Bleeding” and “No Bleeding”. They employ a customized CNN model architecture. It consists of an eight-layer convolutional neural network that is composed of three

convolutional layers (C1-C3), three pooling layers (MP1-MP3) and two fully connected layers (FC1, FC2). Moreover, Support Vector Machine (SVM) [8] classifier is utilized instead of the Softmax layer.

#### B. Bleeding Detection System

1) *Conventional approaches:* The study in [21] extracted color and texture features. These features are used to generate bag of words using K-means clustering algorithm. Next, the Expectation Maximization (EM) is employed on the “Bag-of-Visual-Words” for super-pixel segmentation. From the region of interest, geometric features like centroid, area, and eccentricity are extracted and fed to the SVM classifier [8]. The authors in [22] proposed an approach based on statistical color feature analysis. First, the frame is split into blocks. After that, dark or light blocks are excluded. Moreover, canny operator [23] is applied to discard the edges. Furthermore, Wavelet db2 with soft thresholding [24] is applied to reduce noise. The Red channel of the RGB color space is exploited to detect bleeding regions. More specifically, red ratio is computed for individual pixels. Finally, Support Vector Machine (SVM) is used to classify WCE frames into bleeding and non-bleeding classes. Alternatively, the system described in [25] performs semantic segmentation by classifying the pixels as a “Bleeding” or “No Bleeding” pixel. This results in detecting the bleeding pixel within the frame. More specifically, the proposed system in [26] extracts the Red-Green-Blue (RGB) color feature [26] and the Gray-Level Co-occurrence Matrix (GLCM) texture feature [26]. These two features are combined and fed to Random Tree (RT) [27], Random Forest (RF) [28], and Logistic Model Tree (LMT) [29] classifiers.

2) *Deep learning approaches:* The authors in [30] use AlexNet [31] CNN model to classify the frames as “Bleeding”, or “No Bleeding”. This is a well-known CNN model, which is one of the earliest models that won the ILSVRC run by ImageNet. Once the bleeding frames are separated, they are segmented using SegNet [32] in order to detect the “Bleeding” areas. It is a deep learning model designed for image segmentation. It is constituted of convolutional stacked auto-encoder. Similarly, the authors in [33] use U-Net deep learning segmentation approach to detect “Bleeding” regions in the small intestines. The model architecture has a “U” shape. The model down-samples the input image to a small feature map. Next, it up-samples it. The up-sampling process use skip connections to benefit from the down-sampling process. In fact, at each level, the down-sampled feature map is concatenated to the up-sampled one to generate the next up-sampled feature map. The work in [34] employs a Cascade Proposal network to generate region of interest proposals. These are regions susceptible to include bleeding pattern. The proposed regions are then fed to the Region Proposal Rejection (RPR). The latter is a small network consisting of one convolutional layer, one fully connected layer, and two output layers. It is used to rank the regions based on a score. Its output



is fed to a detection module which predicts the bounding box and the corresponding class. For the testing phase, the unseen image is provided to both a Salient Region Segmentation (SRS) and a Multiregional Region Combination (MRC). While SRS captures the exact location of the regions [34], and MRC that gains adequate coverage of the concerned region and apply the SRS to locate region of interest's positions. Moreover, object boundaries are refined using the Dense Region Fusion (DRF) approach by checking the density of a specific area [34].

### C. Discussion

As it can be noticed, the related works in [7], [9], [12], [14], [18], and [20] classify the frames into “Bleeding” and “No Bleeding”. While the earliest studies in [7] and [9] are based of extracting “hand crafted” features that are fed to a classifier, the works in [12], [14], [18], and [20] exploit deep learning models befitting therefore from the automatic learning of the features. In fact, using deep learning paradigm alleviates the problem of selecting the suitable features which is usually performed through empirical comparison of the features. Nevertheless, classification approaches do not localize the bleeding within the frame. Alternatively, the works in [21], [22], [25], [30], [33], and [34] perform bleeding detection. In particular, the studies in [21], [22], and [25] utilize “hand crafted” features. While the work in [21] and [22] splits the frame into blocks to transform the problem into a set of local problems and identifies in which block the bleeding occurs, the work in [25] perform semantic segmentation through pixelwise classification. The deep learning detection-based approaches in [30] and [33] are segmentation approaches. In fact, they exploit well known deep learning segmentation approaches SegNet and U-Net. Nevertheless, these two approaches are known to be very slow and not suitable for real world applications [35]. On the other hand, the work in [34] is not employing segmentation. It learns a bounding box to localize the anomaly. Specifically, it is based on a customized CNN. Thus, the adopted model could be fit the considered datasets. Moreover, it includes several modules, namely, SRS, MRC, RPR, and detection modules. This is advantageous when compared to end-to-end model. In fact, the error inducted by one of these modules affects all other modules. Moreover, the error of the different modules gets accumulated.

## III. PROPOSED APPROACH

Computer aided-diagnosis can lessen the visualization task and help detecting automatically the MBS. As shown in the related works investigation, MBS aided diagnosis systems are based on image processing and machine learning techniques. In particular, most of the reported works related to detecting MBS employ segmentation techniques. As a result, “hand crafted” features for the segmentation task and for the classification task are required. This can be alleviated by the use of deep learning approaches designed for object detection. Nonetheless, to the best of our knowledge, deep learning models have not been explored for MBS detection. In particular, the end-to-end state of the art YOLO models were not investigated.

YOLO deep learning detection model outperformed the other object detection approaches in many pattern recognition

applications [36], [37]. Moreover, the success of YOLO model and its applicability to real world applications, yield the evolution of the model and the publication of different versions. However, a throughout comparisons of these versions in terms of performance and efficiency needs to be performed. In this regard, YOLO model, specifically, its latest versions YOLOv3 [3], YOLOv4 [4], YOLOv5 [5], and YOLOv7 [6] are investigated for detecting MBS in the GI tract. In the following, we describe the four considered models.

### A. YOLOv3 Architecture

YOLO version 3 (YOLOv3) [3] is an improved version of YOLO which seeks to enhance the performance through the use of residual blocks and different scale feature maps. Inspired by Residual Networks [38] YOLOv3 employs alternatively  $3 \times 3$  and  $1 \times 1$  convolutional layers to form a residual unit. This unit aims at avoiding the vanishing gradient problem faced by very deep network. YOLOv3 is composed of five residual block which incorporate a number of residual units. Since a stride of 2 is used at each residual block, the input is down-sampled five times. In particular, the last three down-sampled feature maps are used for the prediction task. Specifically, after the third residual block, the feature map is down-sampled by factor 8. It is exploited for small object prediction. On the other hand, the output of the fourth residual block is down-sampled by a factor of 16, and it is utilized to generate scale 2 feature map. The latter is employed for medium object prediction. Alternatively, big objects, referred to as scale 1 objects, are predicted using the last residual block for which the feature is down-sampled by a factor of 32. Furthermore, YOLOv3 performs feature fusion to benefit from the feature maps at the different scales. As such, it up-samples scale 1 feature map and concatenate it with scale 2 feature map. The obtained feature map is then up-sampled, and concatenate with scale 3 feature map [38].

### B. YOLOv4 Architecture

YOLOv4 [4] is the fourth version of the YOLO model family. YOLOv4 model architecture is composed of multiple sections. Namely, they are the Input, the Backbone, the Neck, and the Head (dense prediction, and the sparse prediction). The backbone and the neck sections are responsible for feature extraction and aggregation, respectively. In particular, the CNN deep learning model, CSPDarkNet53 [39], is used as a feature extractor in the backbone section. Alternatively, Spatial Pyramid Pooling (SPP) and Path Aggregation Network (PANet) were utilized in the neck section to fuse the features using Bag of Specials (BoS). Finally, the head which is responsible for both localizing the object in the image and classifying it, amounts to YOLOv3 models. It consists of two stage detectors. The first one is the one stage object detector and the second one is the one is the two-stage object detector [4]. Compared to the previous versions of YOLO, YOLOv4 mainly introduced two additional concepts. Bag of Freebies (BoF) and Bag of Specials (BoS). Bag of freebies are a set of techniques that alters the training framework or perform data augmentation. Many techniques can be incorporated for the purpose of enhancing the model performance without affecting on the inference cost [10]. Alternatively, BoS are strategies such as enlarging the receptive field, integrating features, incorporating attention modules, or post-processing.

These strategies aim at significantly enhancing the performance of accuracy at the expense of increasing the inference cost [4].

### C. YOLOv5 Architecture

YOLOv5 [5] is implemented using PyTorch which allows faster training [40]. As such, YOLOv5 allows rapid detection with the same accuracy as YOLOv4. Specifically, YOLOv5 has been proved to have higher performance than YOLOv4 under certain circumstances and partly gained confidence in the computer vision community besides YOLOv4. YOLOv5 model architecture is similar to YOLOv4 architecture. It employs CSPDarknet53 [40] for the backbone section as feature extractor. The latter aims at addressing the gradient in deep networks and decreases the inference time through the use of cross-layer connections between the network's front and back layers. Moreover, it seeks improving the accuracy and utilizing lightweight model. Furthermore, the SPP module referring to the Spatial Pyramid Pooling module, performs maximum pooling with several kernel sizes and then fuses the features by concatenating them together. Additionally, YOLOv5 exploits Path Aggregation Network (PANet) in the neck section as feature aggregator to increase the flow of information and to enhance the object localization. Besides, PANet incorporates a Feature Pyramid Network (FPN) [41]. On the other hand, the head is designed in the same way as YOLOv3 and YOLOv4. Specifically, it produces three different scale feature maps. The CSP network in the backbone is made up from one or more residual units, whereas the CSP network in the neck is made up of new module called CBL modules that replace the residual units. The CBL module consist of Convolution layers, Batch normalization layers, and Leaky ReLU activation function modules [42]. YOLOv5 introduces a new layer referred to as Focus layer [43]. It takes the place of the first three layers of YOLOv3. Therefore, it reduces the GPU requirement and decreases the number of layers.

### D. YOLOv7 Architecture

The most recent YOLO architecture, YOLOv7 [44], is based on YOLOv4 version. The main modifications consist of (i) the introduction of the Extended Efficient Layer Aggregation Network (E-ELAN), (ii) the incorporation of model scaling component, (iii) the use of planned re-parameterized convolution, (iv) the employment of auxiliary head, and (v) the exploration of label assigner mechanism. E-ELAN is a computational component in YOLOv7 backbone part. It enhances the prediction performance continuously by employing “expand, shuffle, merge cardinality”. Alternatively, the model scaling optimizes the number of layers, the number of channels, the number of stages in the feature pyramid, and the resolution of the input image in order to meet the requirements of various problems. Nevertheless, YOLOv7 introduces a new model scaling paradigm which optimizes the scaling factors jointly, not independently one from the other. Similarly, YOLOv7 modifies RepConv by discarding the identity connection. In fact, it uses RepConvN in order to prevent the presence of identity connection for re-parameterized convolution. Moreover, YOLOv7 exploits the Deep Supervision training technique. More specifically, YOLOv7 uses an auxiliary head in the intermediate layers to guide the

training. The head responsible for the final prediction is referred to as lead head. Additionally, to further enhance the training, YOLOv7 outputs soft labels instead of hard one referring to the ground truth.

We propose to compare the performance between different YOLO approaches which are YOLOv3 [3], YOLOv4 [4], YOLOv5 [5], and YOLOv7 [6] in recognizing in recognizing “Bleeding” spots. For this purpose, the considered models need to be trained. Therefore, each YOLO model is fed with images indicating the bleeding areas, if any. Specifically, the coordinates of the bounding boxes surrounding the MBS patterns are provided as input along with the “Bleeding” images. They consist of the upper left corner coordinates (X, Y), the width, and the height of each box. Concerning the “Non-Bleeding” images, no boundary box is specified. To determine the best version of YOLO, the considered YOLO models are evaluated using the test set. Specifically, the different models are tested in terms of the inference time, MBS localization and classification. The best performing model is adopted to build the required system.

## IV. EXPERIMENT

Kvasir-Capsule dataset [45] is considered in this project. It is a dataset of WCE videos collected from clinical examinations performed at the Department of Medicine, Bærum Hospital, and Vestre Viken Hospital Trust in Norway. It consists of 406 “Bleeding” images representing bleeding spots of different size, color, and texture. In addition, it includes 34338 “Non-Bleeding” images representing normal GI tract frames (without bleeding). According to [46], it is not recommended to add images without region of interest (“non-bleeding” images) to the training set. More specifically, “non-bleeding” images should not exceed more than 10% of the total number of images in the training set. As such, only 328 non-bleeding images are first considered. This results in the distribution reported in Table I, where the images are divided into 60% for training, 20% for validation, and 20% testing sets. Nevertheless, in order to get a glimpse of the models’ performance on the real-world, 6000 non-bleeding images are used in the test set. More specifically, both test sets which are the test set after omitting most of non-bleeding images (Test 1) and the test set that containing 6000 background images (Test 2) are assessed.

The available Ground Truth consists of labeling the whole image as including bleeding or not. Nevertheless, in order to train YOLO, a different ground truth should be provided. In fact, the coordinates of the bounding boxes surrounding the bleeding spots should be fed to model to be trained. As such, the dataset is labeled using labeling software tool [47]. As a result, 960 bleeding regions are considered.

TABLE I. DATASET DISTRIBUTION

	<i>Training set</i>	<i>Validation set</i>	<i>Testing set without additional non-bleeding (Test 1)</i>	<i>Testing set with additional non-bleeding (Test 2)</i>
<b>Bleeding</b>	231	75	100	100
<b>Non-bleeding</b>	328	108	86	6000

Two performance measures are considered to evaluate the performance of YOLOv3 [3], YOLOv4 [4], YOLOv5 [5], and YOLOv7 [6] in terms of recognizing MBS. Specifically, we considered Intersection over Union (IoU) [48] and mean Average Precision (mAP) [49], since the localization and the categorization of the object of interest are assessed using these performance measures. Moreover, Floating Point Operations per second (FLOP) [50] is also considered to compare the time efficiency of the considered YOLO models. Fig. 1 shows a comparison between the performances of the considered YOLO models on Test 2 in terms of both mAP and IoU.

As illustrated in Fig. 1, YOLOv3 performs better than YOLOv4 and YOLOv5 in terms of recognition with mAP equal to 0.828. This is an expected outcome since the architecture of YOLOv3 consists of residual blocks. One of them is exploited specifically for small object detections which concord with the small pattern of the bleeding spots. Moreover, in terms of IoU, YOLOv4 achieves an IoU of 0.736 which is better than 0.589 for YOLOv3 and 0.727 for YOLOv5. In fact, YOLOv4 is better in localizing bleeding spots since it incorporates two stage detectors. The first one is called the one stage object detector and the second one is the two-stage object detector. Nevertheless, YOLOv5 exploits path aggregation network that enhances the model localization ability. Alternatively, YOLOv7 achieved the highest IoU and mAP equal to 0.8 and 0.86 respectively. This makes YOLOv7 the most appropriate model to design the proposed approach.

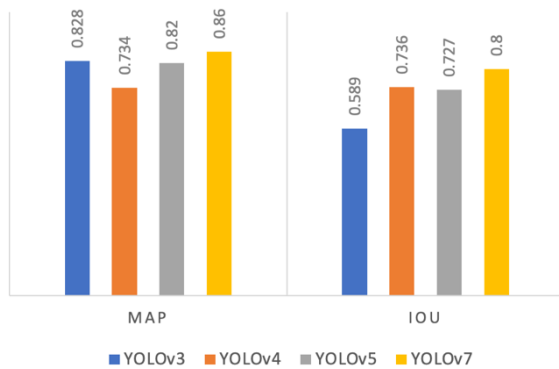


Fig. 1. Performance comparison of YOLOv3, YOLOv4, YOLOv5, and YOLOv7 in terms of mAP and IoU.

Moreover, data augmentation is employed to increase the size of the training data set conveyed to the best performance model, namely YOLOv7 [6]. This is achieved by adding more images to train the model. These images were created by flipping and rotating existing training images. The augmented dataset contains “1056” images. The performance of YOLOv7 without using the augmented data is compared with its performance when training the model with additional data. Table II depicts the performance of YOLOv7 when including and excluding data augmentation. As it can be seen, the augmented dataset improved YOLOv7 performance in terms of mAP.

Furthermore, we compare YOLOv3 [3], YOLOv4 [4], YOLOv5 [5] and YOLOv7 [6] in terms of space complexity. It refers to the space needed to store and train the model. Table III shows the space memory for each model. As depicted,

YOLOv5 requires less space memory due to its optimized implementation, while YOLOv4 needs more space memory.

TABLE II. PERFORMANCE COMPARISON OF YOLOv7 [6] WHEN USING DATA AUGMENTATION AND WITHOUT USING IT

	mAP	IoU	FLOPs
Test results using data augmentation	0.883	0.81	188.9G
Test results without data augmentation	0.86	0.8	188.9G

TABLE III. PERFORMANCE ANALYSIS IN TERMS OF SPACE COMPLEXITY

Model	Space
YOLOv3 Redmon and Farhadi, “YOLOv3.”	123.5 MB
YOLOv4 Bochkovskiy, Wang, and Liao, “YOLOv4.”	491.6 MB
YOLOv5 “Releases • Ultralytics/Yolov5.”	14.4 MB
YOLOv7 Wang, Bochkovskiy, and Liao, “YOLOv7.”	142 MB

As illustrated in in Fig. 1, YOLOv7 exceeds the other models in terms of test result, yet there is no significant increase in term of time complexity. It is noticeable that YOLOv4 consumed more time when training the model. On the other hand, when training YOLOv5 it took the least time, and that is predictable since YOLOv5 uses less floating-point operations. Regarding the time considered to train all four models, Table IV reports the training and testing times per image when using Google Collaboratory to train all models.

TABLE IV. PERFORMANCE ANALYSIS IN TERMS OF TRAINING AND TESTING TIME COMPLEXITY

Model	Training Time (s)	Testing Time (ms)
YOLOv3 Redmon and Farhadi, “YOLOv3.”	6.5295	0.00026
YOLOv4 Bochkovskiy, Wang, and Liao, “YOLOv4.”	23.07	0.00837
YOLOv5 “Releases • Ultralytics/Yolov5.”	3.8103	0.00031
YOLOv7 Wang, Bochkovskiy, and Liao, “YOLOv7.”	11.0554	0.00124

## V. CONCLUSION AND FUTURE WORKS

The arduousness of MBS diagnosis through the burdensome visualization of an eight-hour WCE video of the GI tract has led to the development of aided-diagnosis system. They are based on pattern recognition techniques to detect MBS. In this paper, we proposed to design an aided- diagnosis for MBS detection from WCE video. It is based on deep learning pattern recognition model. In particular, different versions of YOLO model are investigated. Four YOLO models are trained and tested. The comparison and the analysis of the obtained results yielded the selection of the most suitable YOLO model for MBS recognition from WCE videos of the GI tract. Namely, YOLOv7 outperformed the other models.

As future works, the proposed system can be implemented to an applicable and more convenient user-friendly system that

can be used by physicians. Additionally, the performance of the proposed system can be further enhanced by collecting more WCE data to train the model.

## REFERENCES

- [1] T. Wilkins, B. Wheeler, and M. Carpenter, "Upper Gastrointestinal Bleeding in Adults: Evaluation and Management," *Am. Fam. Physician*, vol. 101, no. 5, pp. 294–300, Mar. 2020.
- [2] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2016, pp. 779–788. doi: 10.1109/CVPR.2016.91.
- [3] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement." *arXiv*, Apr. 08, 2018. Accessed: Oct. 16, 2022. [Online]. Available: <http://arxiv.org/abs/1804.02767>
- [4] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection." *arXiv*, Apr. 22, 2020. Accessed: Oct. 16, 2022. [Online]. Available: <http://arxiv.org/abs/2004.10934>
- [5] "Releases • ultralytics/yolov5," GitHub. <https://github.com/ultralytics/yolov5/releases> (accessed Oct. 17, 2022).
- [6] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors." *arXiv*, Jul. 06, 2022. Accessed: Oct. 08, 2022. [Online]. Available: <http://arxiv.org/abs/2207.02696>
- [7] S. Alotaibi, S. Qasim, O. Bchir, and M. M. Ben Ismail, "Empirical Comparison of Visual Descriptors for Multiple Bleeding Spots Recognition in Wireless Capsule Endoscopy Video," in *Computer Analysis of Images and Patterns*, R. Wilson, E. Hancock, A. Bors, and W. Smith, Eds., in *Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer, 2013, pp. 402–407. doi: 10.1007/978-3-642-40246-3\_50.
- [8] Y. Liu and Y. F. Zheng, "One-against-all multi-class SVM classification using reliability measures," in *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.*, Jul. 2005, pp. 849–854 vol. 2. doi: 10.1109/IJCNN.2005.1555963.
- [9] O. Bchir, M. M. Ben Ismail, and N. AlZahrani, "Multiple bleeding detection in wireless capsule endoscopy," *Signal Image Video Process.*, vol. 13, no. 1, pp. 121–126, Feb. 2019, doi: 10.1007/s11760-018-1336-3.
- [10] M. Verma, B. Raman, and S. Murala, "Multi-resolution Local extrema patterns using discrete wavelet transform," in 2014 Seventh International Conference on Contemporary Computing (IC3), Aug. 2014, pp. 577–582. doi: 10.1109/IC3.2014.6897237.
- [11] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: The fuzzy c-means clustering algorithm," *Comput. Geosci.*, vol. 10, no. 2, pp. 191–203, Jan. 1984, doi: 10.1016/0098-3004(84)90020-7.
- [12] R. Shahril, A. Saito, A. Shimizu, and S. Baharun, "Bleeding Classification of Enhanced Wireless Capsule Endoscopy Images using Deep Convolutional Neural Network," p. 18.
- [13] Y. LeCun, L. Bottou, Y. Bengio, and P. Ha, "Gradient-Based Learning Applied to Document Recognition," p. 46, 1998.
- [14] A. Caroppo, A. Leone, and P. Siciliano, "Deep transfer learning approaches for bleeding detection in endoscopy images," *Comput. Med. Imaging Graph.*, vol. 88, p. 101852, Mar. 2021, doi: 10.1016/j.compmedimag.2020.101852.
- [15] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition." *arXiv*, Apr. 10, 2015. Accessed: Oct. 17, 2022. [Online]. Available: <http://arxiv.org/abs/1409.1556>
- [16] S. Mukherjee, "The Annotated ResNet-50," *Medium*, Aug. 18, 2022. <https://towardsdatascience.com/the-annotated-resnet-50-a6c536034758> (accessed Oct. 21, 2022).
- [17] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision." *arXiv*, Dec. 11, 2015. doi: 10.48550/arXiv.1512.00567.
- [18] F. Rustam et al., "Wireless Capsule Endoscopy Bleeding Images Classification Using CNN Based Model," *IEEE Access*, vol. PP, pp. 1–1, Feb. 2021, doi: 10.1109/ACCESS.2021.3061592.
- [19] A. Pujara, "Image Classification With MobileNet," *Analytics Vidhya*, Jul. 15, 2020. <https://medium.com/analytics-vidhya/image-classification-with-mobilenet-cc6fbb2cd470> (accessed Oct. 21, 2022).
- [20] X. Jia and M. Q.-H. Meng, "A deep convolutional neural network for bleeding detection in Wireless Capsule Endoscopy images," in 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, FL, USA: IEEE, Aug. 2016, pp. 639–642. doi: 10.1109/EMBC.2016.7590783.
- [21] P. Sivakumar and B. M. Kumar, "A novel method to detect bleeding frame and region in wireless capsule endoscopy video," *Clust. Comput.*, vol. 22, no. S5, pp. 12219–12225, Sep. 2019, doi: 10.1007/s10586-017-1584-y.
- [22] S. Suman et al., "Detection and Classification of Bleeding Region in WCE Images using Color Feature." 2017. doi: 10.1145/3095713.3095731.
- [23] F. Wu, C. Zhu, J. Xu, M. W. Bhatt, and A. Sharma, "Research on image text recognition based on canny edge detection algorithm and k-means algorithm," *Int. J. Syst. Assur. Eng. Manag.*, vol. 13, no. S1, pp. 72–80, Mar. 2022, doi: 10.1007/s13198-021-01262-0.
- [24] P. V. V. Kishore, A. S. C. S. Sastry, A. Kartheek, and Sk. H. Mahatha, "Block based thresholding in wavelet domain for denoising ultrasound medical images," in 2015 International Conference on Signal Processing and Communication Engineering Systems, Guntur, India: IEEE, Jan. 2015, pp. 265–269. doi: 10.1109/SPACES.2015.7058262.
- [25] K. Pogorelov et al., "Bleeding detection in wireless capsule endoscopy videos — Color versus texture features," *J. Appl. Clin. Med. Phys.*, vol. 20, no. 8, pp. 141–154, 2019, doi: 10.1002/acm2.12662.
- [26] C. Sri Kusuma Aditya, M. Hani'ah, R. R. Bintana, and N. Suciati, "Batik classification using neural network with gray level co-occurrence matrix and statistical color feature extraction," in 2015 International Conference on Information & Communication Technology and Systems (ICTS), Surabaya: IEEE, Sep. 2015, pp. 163–168. doi: 10.1109/ICTS.2015.7379892.
- [27] S. Kalmegh, "Analysis of WEKA Data Mining Algorithm REPTree, Simple Cart and RandomTree for Classification of Indian News," vol. 2, no. 2, p. 9.
- [28] E. K. Sahin, I. Colkesen, and T. Kavzoglu, "A comparative assessment of canonical correlation forest, random forest, rotation forest and logistic regression methods for landslide susceptibility mapping," *Geocarto Int.*, vol. 35, no. 4, pp. 341–363, Mar. 2020, doi: 10.1080/10106049.2018.1516248.
- [29] M. Abedini, B. Ghasemian, A. Shirzadi, and D. T. Bui, "A comparative study of support vector machine and logistic model tree classifiers for shallow landslide susceptibility modeling," *Environ. Earth Sci.*, vol. 78, no. 18, p. 560, Sep. 2019, doi: 10.1007/s12665-019-8562-z.
- [30] T. Ghosh and J. Chakareski, "Deep Transfer Learning for Automated Intestinal Bleeding Detection in Capsule Endoscopy Imaging," *J. Digit. Imaging*, vol. 34, no. 2, pp. 404–417, Apr. 2021, doi: 10.1007/s10278-021-00428-3.
- [31] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.
- [32] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation." *arXiv*, Oct. 10, 2016. Accessed: Oct. 21, 2022. [Online]. Available: <http://arxiv.org/abs/1511.00561>
- [33] P. Coelho, A. Pereira, A. Leite, M. Salgado, and A. Cunha, "A Deep Learning Approach for Red Lesions Detection in Video Capsule Endoscopies," in *Image Analysis and Recognition*, A. Campilho, F. Karray, and B. ter Haar Romeny, Eds., in *Lecture Notes in Computer Science*, vol. 10882. Cham: Springer International Publishing, 2018, pp. 553–561. doi: 10.1007/978-3-319-93000-8\_63.
- [34] L. Lan, C. Ye, C. Wang, and S. Zhou, "Deep Convolutional Neural Networks for WCE Abnormality Detection: CNN Architecture, Region Proposal and Transfer Learning," *IEEE Access*, vol. 7, pp. 30017–30032, 2019, doi: 10.1109/ACCESS.2019.2901568.
- [35] M. Z. Alom, C. Yakopcic, M. Hasan, T. M. Taha, and V. K. Asari, "Recurrent residual U-Net for medical image segmentation," *J. Med. Imaging*, vol. 6, no. 01, p. 1, Mar. 2019, doi: 10.1117/1.JMI.6.1.014006.

- [36] T. Diwan, G. Anirudh, and J. V. Tembhurne, "Object detection using YOLO: challenges, architectural successors, datasets and applications," *Multimed. Tools Appl.*, Aug. 2022, doi: 10.1007/s11042-022-13644-y.
- [37] "(9) (PDF) KVASIR: A Multi-Class Image Dataset for Computer Aided Gastrointestinal Disease Detection." [https://www.researchgate.net/publication/316215961\\_KVASIR\\_A\\_Multi-Class\\_Image\\_Dataset\\_for\\_Computer\\_Aided\\_Gastrointestinal\\_Disease\\_Detection](https://www.researchgate.net/publication/316215961_KVASIR_A_Multi-Class_Image_Dataset_for_Computer_Aided_Gastrointestinal_Disease_Detection) (accessed Oct. 31, 2022).
- [38] Ju, Luo, Wang, Hui, and Chang, "The Application of Improved YOLO V3 in Multi-Scale Target Detection," *Appl. Sci.*, vol. 9, no. 18, p. 3775, Sep. 2019, doi: 10.3390/app9183775.
- [39] N. Kwak and D. Kim, "Object detection technology trend and development direction using deep learning," *Int. J. Adv. Cult. Technol.*, vol. 8, no. 4, pp. 119–128, Dec. 2020, doi: 10.17703/IJACT.2020.8.4.119.
- [40] M. Sozzi, S. Cantalamessa, A. Cogato, A. Kayad, and F. Marinello, "Automatic Bunch Detection in White Grape Varieties Using YOLOv3, YOLOv4, and YOLOv5 Deep Learning Algorithms," *Agronomy*, vol. 12, no. 2, Art. no. 2, Feb. 2022, doi: 10.3390/agronomy12020319.
- [41] T.-K. Nguyen, L. Vu, V. Vu, T.-D. Hoang, S.-H. Liang, and M.-Q. Tran, "Analysis of Object Detection Models on Duckietown Robot Based on YOLOv5 Architectures," vol. 4, pp. 17–12, Mar. 2022.
- [42] X. Xu, X. Zhang, and T. Zhang, "Lite-YOLOv5: A Lightweight Deep Learning Detector for On-Board Ship Detection in Large-Scene Sentinel-1 SAR Images," *Remote Sens.*, vol. 14, no. 4, p. 1018, Feb. 2022, doi: 10.3390/rs14041018.
- [43] K. Patel, C. Bhatt, and P. L. Mazzeo, "Deep Learning-Based Automatic Detection of Ships: An Experimental Study Using Satellite Images," *J. Imaging*, vol. 8, no. 7, p. 182, Jun. 2022, doi: 10.3390/jimaging8070182.
- [44] G. Boesch, "YOLOv7: The Most Powerful Object Detection Algorithm (2022 Guide)," *viso.ai*, Aug. 11, 2022. <https://viso.ai/deep-learning/yolov7-guide/> (accessed Oct. 08, 2022).
- [45] P. H. Smedsrud et al., "Kvasir-Capsule, a video capsule endoscopy dataset," *Sci. Data*, vol. 8, no. 1, Art. no. 1, May 2021, doi: 10.1038/s41597-021-00920-z.
- [46] "how to use Background images in training? • Issue #2844 • ultralytics/yolov5." <https://github.com/ultralytics/yolov5/issues/2844> (accessed Feb. 07, 2023).
- [47] "heartexlabs/labelImg." *Heartex*, Oct. 30, 2022. Accessed: Oct. 30, 2022. [Online]. Available: <https://github.com/heartexlabs/labelImg>
- [48] Naoki, "Object Detection: Intersection over Union (IoU)," *Medium*, Oct. 08, 2022. <https://naokishibuya.medium.com/object-detection-intersection-over-union-iou-f7b91555eb5f> (accessed Oct. 26, 2022).
- [49] B. Wang, "A Parallel Implementation of Computing Mean Average Precision." *arXiv*, Jun. 19, 2022. Accessed: Oct. 22, 2022. [Online]. Available: <http://arxiv.org/abs/2206.09504>
- [50] "Floating-Point Operation - an overview | ScienceDirect Topics." <https://www.sciencedirect.com/topics/computer-science/floating-point-operation> (accessed Oct. 30, 2022).

# A Novel Feature Fusion for the Classification of Histopathological Carcinoma Images

Salini S Nair<sup>1</sup>, M. Subaji<sup>2</sup>

School of Computer Science and Engineering, Vellore Institute of Technology (VIT), Vellore, India<sup>1</sup>  
Institute for Industry and International Programmes, Vellore Institute of Technology, Vellore, India<sup>2</sup>

**Abstract**—Breast cancer is a significant global health concern, demanding advanced diagnostic approaches. Although traditional imaging and manual examinations are common, the potential of artificial intelligence (AI) and machine learning (ML) in breast cancer detection remains underexplored. This study proposes a hybrid approach combining image processing and ML methods to address breast cancer diagnosis challenges. The method utilizes feature fusion with gray-level co-occurrence matrix (GLCM), local binary patterns (LBP), and histogram features, alongside an ensemble learning technique for improved classification. Results demonstrate the approach's effectiveness in accurately classifying three carcinoma classes (ductal, lobular, and papillary). The Voting Classifier, an ensemble learning model, achieves the highest accuracy, precision, recall, and F1-scores across carcinoma classes. By harnessing feature extraction and ensemble learning, the proposed approach offers advantages such as early detection, improved accuracy, personalized medicine recommendations, and efficient analysis. Integration of AI and ML in breast cancer diagnosis shows promise for enhancing accuracy, effectiveness, and personalized patient care, supporting informed decision-making by healthcare professionals. Future research and technological advancements can refine AI-ML algorithms, contributing to earlier detection, better treatment outcomes, and higher survival rates for breast cancer patients. Validation and scalability studies are needed to confirm the effectiveness of the proposed hybrid approach. In conclusion, leveraging AI and ML techniques has the potential to revolutionize breast cancer diagnosis, leading to more accurate and personalized detection and treatment. Technology-driven advances can significantly impact breast cancer care and management.

**Keywords**—Breast cancer; machine learning; artificial intelligence; feature extraction; ensemble classifier

## I. INTRODUCTION

Cancer is one of the complex and devastating diseases that continues to pose significant challenges to global healthcare systems and individuals worldwide [1]. It is one of the most dreadful diseases that is not easily curable. In the body, aberrant cells develop and spread out of control, which is a term used to describe a set of disorders called Cancer [2]. In the modern world methods like computerized tomography (CT) scans, magnetic resonance imaging (MRI) scans, positron emission tomography (PET) scans, etc. are used to detect this disease [3]. Breast cancer stands out among the numerous types of cancer as one of the commonest and worrisome forms, impacting millions of people every year [4]. Breast cancer often affects the breast tissue and frequently begins in the milk-producing glands (lobules) or the ducts that

supply milk to the nipple. Mammography, a low-dose X-ray examination of the breast, is the most common technique used for the common detection of breast abnormalities [5]. In addition to this clinical breast examination is another method performed by professionals to detect it. The application of artificial intelligence and associated approaches is still not well practiced for this goal, despite the fact that there are numerous computerized automated procedures utilized for the diagnosis and detection of breast cancer [6]. Manual work that has to be done to diagnose and detect even after this automated process is still cumbersome since it demands intelligent decision-making [7]. The introduction of AI-ML on it will be the solution to it, where doctors do not need to manually examine and diagnose the disease [8]. The advancement of technology has enabled different kinds of methods to detect cancer which mainly include Liquid biopsy, Genome profiling, Image techniques, Metabolomics, Optical techniques, and finally AI -ML techniques [9]. Although these cutting-edge techniques have substantially improved cancer detection, their use may differ depending on the type and stage of the disease, the accessibility of resources, and the state of the healthcare system. Moving forward AI-ML techniques have the potential to be further honed and improved by future research and technological developments, which could ultimately result in earlier cancer diagnosis, better treatment outcomes, and higher overall survival rates for cancer patients [10]. Here we use traditional image processing and machine learning techniques in a hybrid way to realize the detection module. The major advantages of using this technique are early detection, improved accuracy, helping to suggest better personalized medicine, faster and more efficient analysis, integration and multimodal data, and continuous learning and improvement [11]. The benefits of utilizing AI-ML algorithms for cancer diagnosis, as described above, are generally very applicable to the particular situation of breast cancer.

The proposed approach is driven by a comprehensive set of motivations and potential benefits that promise to significantly advance the field of histopathological image classification. Its core aim is to elevate the accuracy and robustness of this critical task. To achieve this, the approach combines three distinct feature extraction techniques: GLCM, which captures pixel-level spatial relationships; LBP, designed to characterize intricate texture patterns; and histogram features, which provide a global view of intensity distribution within the images. By amalgamating these diverse features, the approach seeks to create a holistic representation of the carcinoma images, enabling the model to capture both local nuances and global context, thus enhancing classification

accuracy. Histopathological carcinoma images are notoriously diverse due to variations in tissue preparation, staining, and imaging conditions. Therefore, another vital motivation is to bolster the model's resilience to such variability. The fusion of GLCM, LBP, and histogram features offers a multi-faceted approach to understanding these images, making it more adaptable to different staining protocols and equipment, ultimately resulting in a more reliable diagnostic tool.

Moreover, the approach combats overfitting—a common challenge in machine learning—by employing an ensemble of classifiers. Ensemble methods aggregate the decisions of multiple classifiers, reducing the risk of the model memorizing noise in the training data and improving its generalization performance. This becomes crucial in histopathological image classification, where datasets can be limited in size and prone to noise. Class imbalance is yet another challenge in this domain, with some carcinoma subtypes having fewer samples than others. The fusion technique, coupled with appropriate strategies like weighted voting, can help address these class imbalance issues, ensuring that the model's performance is not skewed towards the majority class, which can be critical for effective clinical diagnosis.

The combined use of different feature types also enhances the interpretability of classification results. Researchers and clinicians can gain insights into which aspects of the images are most influential in making the classification decisions. This not only provides transparency in the model's decision-making process but also aids in building trust in its recommendations. Additionally, the versatility of this approach extends to its potential for transferability. By fusing diverse features and leveraging ensemble classifiers, it can potentially be applied to related image classification tasks within the medical domain, paving the way for broader applicability and impact.

In essence, the given method is a forward-thinking approach that aims to improve classification accuracy, increase model robustness, and enhance the overall performance of histopathological carcinoma image classification. By integrating multiple feature extraction methods and harnessing the power of ensemble classifiers, this approach holds great promise in delivering more accurate and reliable cancer diagnoses, thus contributing significantly to the field of medical image analysis and ultimately benefiting patients and healthcare providers.

These techniques could enhance the precision, effectiveness, and personalization of breast cancer diagnosis and treatment, improving patient outcomes and assisting doctors in their decision-making [12]. In this proposed work we have used a feature fusion for extracting robust features and ensemble learning [13] for better classification performance on classifying the three classes of carcinoma images said ductal, lobular, and papillary. We have used features like GLCM, LBP [14] and Histogram [15]. The novel approach outperformed existing techniques even without using any computational heavy deep learning technique. The remaining portion of the paper is described as essential

preliminaries, detailed methodology, obtained results analysis and discussions, conclusion and the future work.

## II. LITERATURE REVIEW

Alqudah et al. [16] proposed a new sliding window technique for local feature extraction from 25 sliding windows for each image. They used the LBP for features extraction of each window, support vector machine (SVM) to classify the windows, and to find the final class based on the majority voting technique. For the categorization of breast cancer, Gour et al. [17] introduced ResHist, a 152-layered convolutional neural network based on residual learning. They extracted discriminative features from the histopathological images and used the data augmentation technique to enhance the model's performance. Gandomkar et al. [18] have proposed classifying hematoxylin-eosin stained breast digital slides of 81 patients resulting in 7786 images in all. They demonstrated a system known as MuDeRN, which stands for "MULTI-category classification of breast histopathological image using DEep Residual Networks." A deep residual network (ResNet) of 152 layers has been trained to categorize patches from the images in the first stage of the project, which comprises of two stages. Second, the images classified as malignant and benign were classified into four subtypes. Using a meta-decision tree, the authors combined the outputs of ResNet's processed images in different magnification factors. Multiscale generalized radial basis function (MSRBF) neural networks were recommended by Beltran-Perez et al. [19] for the extraction and categorization of image features. Three steps make up the architecture described in this work: first, an input-output model is derived from the image; second, high-level image features are extracted from the model; and third, a module for classification is intended to forecast breast cancer. An approach based on deep convolutional neural network that supports 16 layers (VGGNet-16) has been proposed by Kumar et al. [20] who also assessed how well the fused framework performed in comparison to other classifiers like the support vector machine and random forest. They increase the data size using data augmentation.

Li et al. [21] have evaluated histological images using convolutional neural network (CNN) architecture for classification. In order to improve feature information, authors proposed densely-connected-convolutional network (DenseNet) as the fundamental building block and interspersed it with the squeeze-and-excitation network (SENet) module. Vo et al. [22] have proposed data augmentation approaches to improve classification performance in addition to increasing the diagnosis effectiveness of biopsy tissue utilizing hematoxylin and eosin-stained images. To improve classification performance in the situation of a small number of breast cancer images and imbalanced training data, they have presented an ensemble of deep convolutional neural networks (DCNNs) trained to extract visual features from multiscale images and used gradient boosting tree classifiers. Whereas Saxena et al. [23] have proposed a hybrid ML model to solve the class imbalance problem. They created the kernelized weighted extreme learning machine and the pre-trained ResNet50 for breast cancer classification using histological image. Alom et al. [24] state that the Inception Recurrent Residual

Convolutional Neural Network (IRRCNN) is assessed for breast cancer classification at the image, patient, and patch levels. Boumaraf et al. [25] have put forward the deep neural network ResNet-18, and transfer learning helps to avoid overfitting and boost the training speed on histopathological images. Furthermore, they used global contrast normalization (GCN) to strengthen the approach and three-fold data augmentation to enhance the model. On histopathology images, Burçak et al.'s deep convolutional neural network [26] presents a method for automatically identifying and categorizing malignant areas. For quicker backpropagation learning, they computed the network's starting weight and updated the model parameters using a variety of algorithms, including stochastic gradient descent (SGD), nesterov accelerated gradient (NAG), adaptive gradient (AdaGrad), root mean squared propagation (RMSprop), AdaDelta, and Adam. A graphics processing unit with compute unified device architecture (CUDA) support is utilized in parallel computing architecture for quick processing. Xie et al. [27] investigated to extract expressive features from images of breast cancer's histopathology. They suggested Inception\_V3 and Inception\_ResNet\_V2 deep convolutional neural networks that have been developed using transfer learning strategies. Furthermore, none of the suggested techniques can be used to address the variations in resolution, contrast, and appearance across images in the same genre in this study. Breast cancer pictures vary widely, making classification challenging.

Jiang et al. [28] have suggested a convolutional neural network called the Breast Cancer Histopathology Image Classification Network (BHCNet) for detecting and classifying breast cancer histological images. Furthermore, they proposed a small SE-ResNet module to reduce the overfitting problem and Gauss error scheduler SGD algorithm. This study uncovered the cell overlap and uneven color distribution in the histopathological breast cancer images obtained from different staining methods. To address the imbalanced class problem, Han et al. [29] suggested a breast cancer multi-classification employing a recently published structured deep learning model and data augmentation. Kumar et al. [30] proposed the contrast-limited adaptive histogram equalization approach to enhance microscopic biopsy images, and for segmentation,  $k$ -means clustering is used. Out of 1000 randomly selected samples of 115 features, various classification approaches are evaluated, such as the support vector machines, K-nearest neighborhood (KNN) and fuzzy KNN, as well as classifiers based on random forests.

Sheikh et al. [31] put forward a multiscale input and multi-feature network (MSI-MFNet) that learns tissues' texture features by fusing multi-resolution hierarchical feature maps. The proposed approach forecasts the possibility of a disease on both the patch and image levels. Using the structural and statistical data from the images, Nahid et al. [32] proposed novel deep neural network (DNN) approaches. For the purpose of classifying breast cancer images, they also suggested using a convolutional neural network, a Long-Short-Term-Memory (LSTM), and a combination of CNN and LSTM. Once they had extracted the features from the novel

DNN model, they used Softmax and SVM layers to make decisions. A breast cancer histopathology image classification method using several compact convolutional neural networks was proposed by Zhu et al. [33]. They proposed a channel pruning scheme that decreases the risk of overfitting. The different data partition and composition-based models were assembled to enhance the model's ability to classify the data. The graph convolutional network developed by Gong et al. [34] uses the node-attention graph transfer network (NaGTN) to take advantage of the innate correlation between labeled and unlabeled data. In order to undertake the extraction of knowledge for the target domain, this approach uses a fully labeled source domain. Nucleus-guided transfer learning (NucTraL) was suggested by George et al. [35] as a technique for classifying breast tumors. Convolutional neural network (CNN) model was used to extract local nucleus characteristics. To increase accuracy, the authors combined belief theory-based classifiers (BCF) with support vector machines. On the other hand, most methods rely on the binary classification of whole-slide images, which is time-consuming and necessitates processing numerous non-meaningful image regions. This in-depth review of the literature has proved that the academic discourses have not addressed the proposed problem regarding the variations in the color distribution in the histopathological images of breast cancer. Most of the methods address only binary classification problems.

### III. PRELIMINARIES

#### A. Gray-Level Co-Occurrence Matrix

Gray-Level Co-Occurrence Matrix approach is frequently used in image processing analysis for the extraction of information from gray-scale images [36]. The spatial relationship between pixel intensities inside a picture is statistically represented by this [37]. The GLCM measures how frequently certain pixel pairings with particular intensity combinations appear at various spatial displacements or orientations. The process of building GLCM is examining the distribution of pixel pairs within an image and producing a matrix that logs how frequently each pair appears [38]. Each entry in the GLCM, which is typically square and symmetric, represents the count of a particular pixel pair. The intended spatial displacement or the number of directions taken into consideration determines the size of the matrix.

#### B. Local Binary Patterns

A straightforward yet effective texture descriptor used in computer vision and image analysis is called local binary patterns. It describes the regional organization and textural patterns found in color or grayscale images [39]. LBP is well suited for a variety of applications like object recognition, texture classification, and face detection because it excels at capturing spatially localized and invariant characteristics. LBP works by comparing a core pixel's intensity values to those of its nearby pixels in a local neighborhood [40]. The effectiveness of LBP's computations is one of its benefits. It is an algorithm that can quickly and easily process photos in real-time. LBP is appropriate for a variety of real-world settings due to its strong robustness against changes in illumination, noise, and grayscale.



### C. Histogram Features

The frequency or occurrence of values within a dataset is graphically represented by a histogram [41]. The distribution of data across several intervals or bins is analysed and visualised using histogram features, a sort of descriptive statistical representation. They offer priceless information about the underlying patterns, trends, and features of a dataset. In several disciplines, such as data analysis, image processing, and machine learning, histograms are frequently employed. Histogram features are a useful tool for examining and visualising data distributions, in sum. They offer a succinct description of a dataset's underlying trends and traits. Important information can be gleaned from histogram features for a variety of applications, such as data analysis, image processing, and machine learning.

## IV. METHODOLOGY

This section describes the methodology of the proposed work. The Fig. 1 depicts overall architecture which consists of several sub-stages.

### A. Dataset

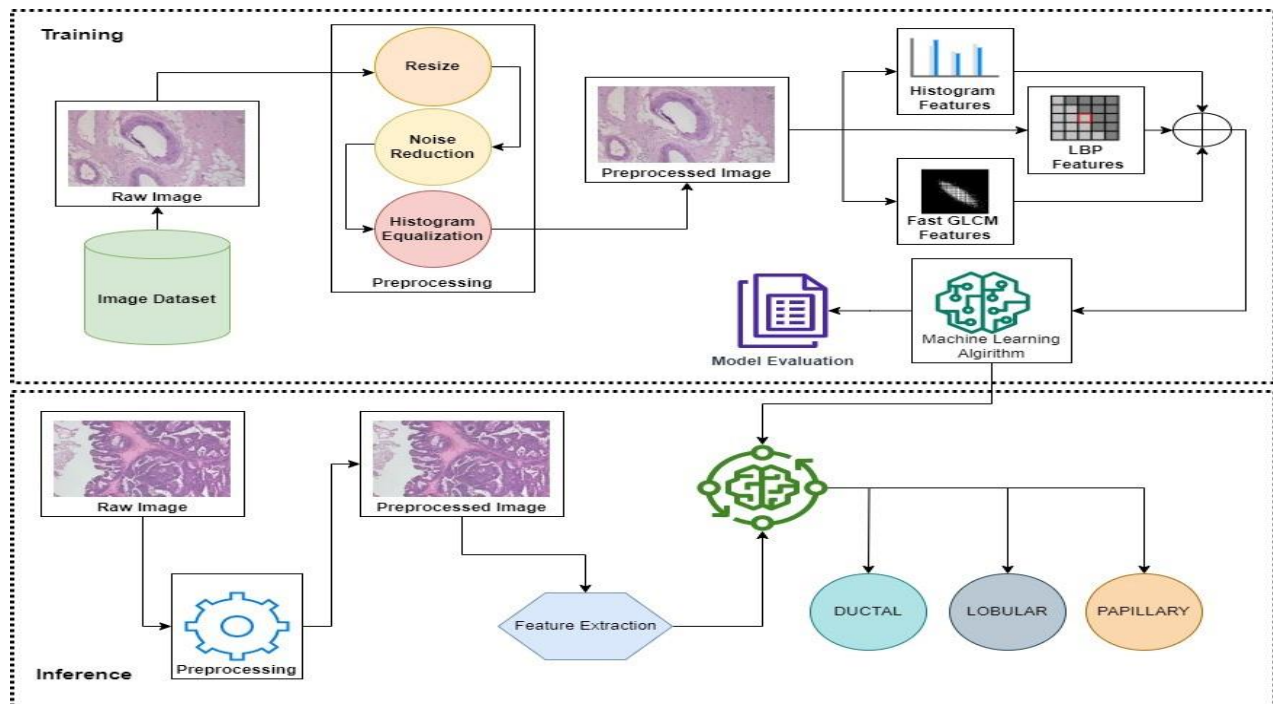


Fig. 1. Architecture.

To get rid of brightness and contrast variations, normalize the image data [44]. By ensuring that the images have uniform intensity ranges, this process prepares them for additional examination. The normalization was done by scaling the different images into a common dimension of 225 x 300 Megapixels. Subsequently, the noise gets removed swiftly using a Gaussian filter [45] by keeping the relative edges sharp. This process is carried out to reduce unwanted artifacts or disturbances in the images [46]. Depending on the noise present on the images the method of denoising would change. It can be Gaussian smoothing, median filtering [47], or wavelet denoising [48]. The final process is histogram

In the present study, we have used images from the Breast Cancer Histopathological (BreakHis) dataset, which includes 7909 images of 82 patients at four different magnifications [42]. The data set was gathered from Brazil's Pathological Anatomy and Cytopathology (P&D) Lab. Images have a dimension of 700 460 pixels, are in the PNG (Portable Network Graphics) format, and are in the 3-channel RGB (Red-Green-Blue) format with an 8-bit depth per channel. Hematoxylin and eosin were used to stain the biopsy samples for breast tissue on slides. The dataset contains information on malignant and benign breast tumors. We have only considered a subset consisting of malignant tumors in the BreakHis dataset. The subset chosen is based on the suggestions of domain experts. It contains the most commonly found cancer patterns from different patients.

### B. Pre-processing

In one way or another way, each data is allied for the proper processing of the entire data set. Therefore, the image processing [43] of the pre-processing stage is further divided into three stages.

equalization [49] to boost the contrast of the image and the visibility of key details. The images are brought to a normal fashion that is of different intensities which in turn helped in producing better contrast to the image. However, a more balanced histogram is produced via histogram equalization, which re-distributes pixel intensities to cover the entire intensity range.

### C. Feature Engineering

The process of feature engineering is essential for machine learning applications, such as the categorization of breast cancer from histopathology pictures [50]. To enhance a

prediction model's performance, important and instructive elements from raw data must be extracted. Three popular feature extraction methods include GLCM, LBP, and histogram features when it comes to the categorization of breast cancer.

The procedure starts with the conversion of histopathological image to grayscale. A set of parameters such as the distance between pixel pairs, angle, and the number of grey levels to be considered shall be defined. Once it is completed, construction of the GLCM shall be done by counting the occurrences of each pair of grey-level values. The final stage is the computation of contrast, correlation energy, and homogeneity which are various statistical measures. Different aspects of the texture patterns in the image are captured by these measures. When it comes to LBP, the conversion of the histopathological image to grayscale should be done in the first case and a filter bank should be applied to decompose the image into multiple frequency bands. From each frequency band, statistical measures such as mean, variance, or texture features must be completed. Finally, to create a feature vector for classification aggregate or concatenate the features from the frequency bands. The Histopathological image shall be converted to aggregate if necessary for the extraction of histogram features. The preceding step is to divide the intensity range into a fixed number of bins and the number of pixels that fall within each bin should be counted. Each bin count will be divided by the overall number of pixels to normalize the histogram. Later, mean, variance or skewness can be calculated if required. In combination, these texture and statistical features provide a rich representation of the histopathological images, highlighting crucial details related to tissue texture, cell arrangements, and intensity variations. Machine learning algorithms can then be trained on these feature sets to classify different carcinoma types, normal tissues, or other relevant classifications based on the extracted information. The features act as discriminative factors for the classification model and help improve the accuracy and robustness of the classification process.

#### D. Ensemble Classifier

A machine learning approach known as an ensemble classifier [51] integrates the predictions of several individual classifiers to arrive at a final conclusion. It is especially helpful for tackling complex issues where a single classifier

would not produce good outcomes. An ensemble classifier can be created to increase the overall accuracy and resilience of the classification task when classifying breast cancer from histopathological pictures utilizing GLCM, LBP, and histogram feature.

Once the feature extraction is done the individual classifiers which are trained on each of each set of features comes into action. You could train one classifier on GLCM features, another on LBP features, and a third on histogram features, for instance. Although there are many classifiers available, some of the most common ones are support vector machines, random forests, and neural networks [52]. The next step is construction of ensemble. Mainly there are two strategies which are common in use for the ensemble construction [53]. Voting is the first one. Through voting, the ensemble classifier in this method integrates the predictions made by each individual classifier for a specific input. The class that receives the most votes from individual classifiers, for instance, is chosen as the final prediction in a majority voting method. The second strategy is weighted averaging. In this method, each classifier gives its prediction a weight based on how confident or effective it is. Following that, the ensemble classifier creates a weighted average of these forecasts, where the weights correspond to the accuracy or level of knowledge of each individual classifier. Once all these are done prediction and decision making are the last steps of classification. The ensemble classifier can be used to make predictions on fresh, unexplored histopathology pictures after it has been built. Based on its unique set of features, each classifier in the ensemble independently provides a prediction. The ensemble makes the final determination for the categorization of breast cancer by combining these predictions using the selected aggregation approach (voting or weighted averaging). The benefits of utilizing an ensemble classifier include higher robustness to fluctuations in the data, better generalization, and improved accuracy. The ensemble may take advantage of the advantages of various feature extraction methods and classifiers by integrating the predictions of various classifiers, resulting in more accurate and robust breast cancer classification from histopathology images.

#### V. RESULT ANALYSIS AND DISCUSSION

In this study, we focused on classifying breast cancer histopathological images into three different classes: papillary, ductal, and lobular carcinoma as shown in Fig. 2.

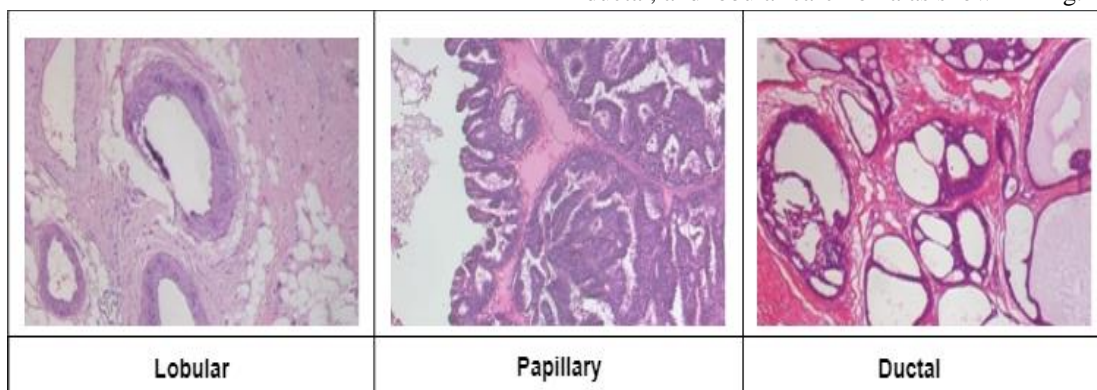


Fig. 2. Carcinoma classes.

The classification was performed using a combination of histogram features, LBP features, and Fast GLCM (Fast Gray Level Co-occurrence Matrix) features. Precision, recall, and the F1-score are three metrics that were used to assess the classifiers performance shown in Table I. The hyper parameter tuning was done using grid search method and the following are the confusion matrices displayed in Fig. 3.

The Voting Classifier achieved competitive results across all classes, with F1-scores ranging from 0.87 for ductal carcinoma to 0.94 for lobular carcinoma. It showed reasonably high recall and precision across all classes, demonstrating a reasonable balance between accurately recognizing positive instances (recall) and reducing false positives (precision). Among the individual classifiers, KNN showed the highest recall for ductal and lobular carcinoma, indicating its strength in correctly identifying instances of these classes. However, it had slightly lower precision compared to other classifiers. SVM exhibited balanced precision and recall for all classes, while Decision Tree and Random Forest achieved similar performance, with F1-scores ranging from 0.82 to 0.87. Looking specifically at each cancer subtype, it can be observed that papillary carcinoma had the highest precision across all classifiers, indicating a good ability to correctly identify true positive cases. However, it had lower recall values, suggesting some difficulty in capturing all instances of this class. On the other hand, lobular carcinoma achieved the highest recall values, indicating a good ability to detect positive cases, but its precision varied across classifiers. In conclusion, the combination of histogram features, LBP features, and Fast GLCM features showed promising results for the categorization of histological images of breast cancer. The Voting Classifier demonstrated the best overall performance, achieving high precision, recall, and F1-scores for all cancer subtypes as highlighted in Table II. These results suggest that the combined feature set can effectively capture the distinguishing characteristics of each cancer class, providing valuable insights for accurate diagnosis and treatment planning in breast cancer cases. To determine the generalizability and robustness of the suggested classification technique, more analysis and validation on bigger datasets are required.

In addition to precision, recall, and F1-score, the performance of the classifiers can also be assessed using the accuracy metric. Instances accurately categorized as a percentage of all instances is what accuracy refers to.

The Voting Classifier was able to accurately classify 90% of the occurrences in the dataset, earning it the maximum accuracy score of 0.90. This classifier outperformed the individual classifiers and demonstrated the best overall

performance. Random Forest also performed well with an accuracy of 0.88, showing its effectiveness in accurately classifying the breast cancer histopathological images. SVM achieved an accuracy of 0.86, indicating a relatively high level of accuracy in its predictions. KNN showed an accuracy of 0.82, while Decision Tree had the lowest accuracy of 0.75 among the classifiers considered. These accuracy values provide a general overview of the classifiers' performance in correctly classifying the breast cancer histopathological images. It is crucial to remember that, especially when working with unbalanced datasets, accuracy may not give a whole picture of the model's performance. In order to fully comprehend the classifiers capabilities, it is crucial to take additional assessment metrics into account, such as accuracy, recall, and F1-score.

TABLE I. EVALUATION METRICS

Class	Voting Classifier			KNN			SVM			Decision Tree			Random Forest		
	Precision	recall	f1-score	Precision	recall	f1-score	Precision	recall	f1-score	Precision	recall	f1-score	Precision	recall	f1-score
Ductal	0.83	<b>0.92</b>	0.87	0.71	0.96	0.82	0.85	0.88	0.87	0.83	0.73	0.78	0.83	0.92	0.87
Lobular	0.89	<b>1.00</b>	0.94	0.84	0.94	0.89	0.85	1.00	0.92	0.75	0.91	0.82	1.00	0.94	0.89
Papillary	0.93	<b>0.92</b>	0.95	0.82	0.81	0.80	0.89	0.86	0.86	0.70	0.62	0.66	0.93	0.74	0.82

TABLE II. TEST ACCURACY

Classifier	Random Forest	KNN	SVM	Decision tree	Voting Classifier
Accuracy	0.88	0.82	0.86	0.75	<b>0.90</b>

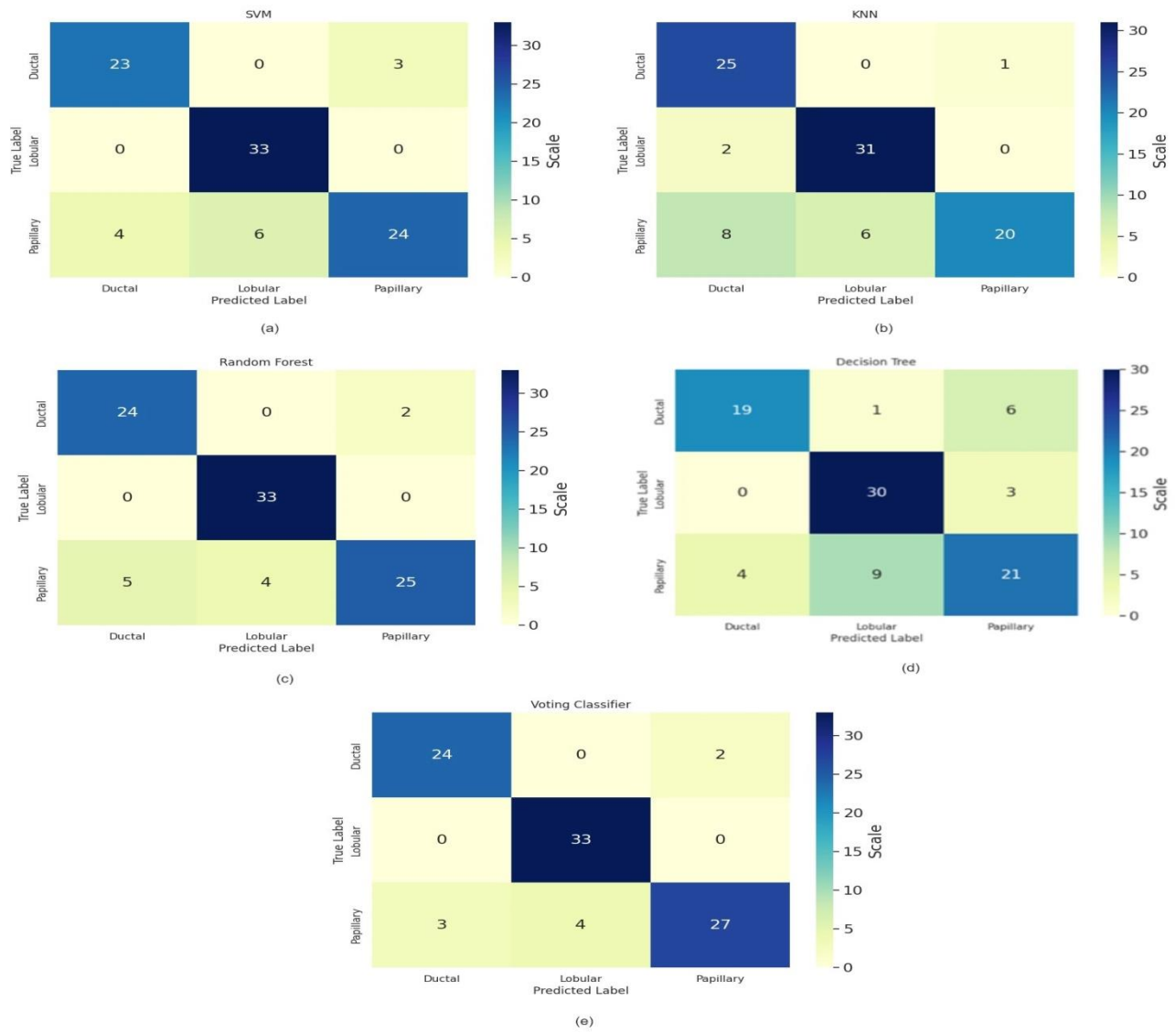


Fig. 3. Confusion matrix of different classifiers.

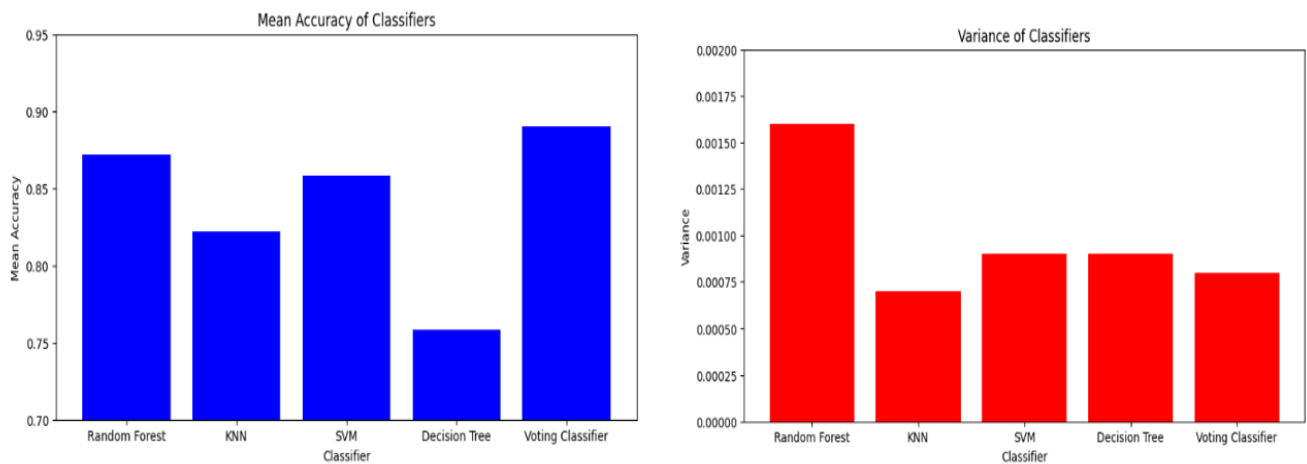


Fig. 4. Mean and variance cross validation accuracy of classifiers.

Overall, the Voting Classifier showed the highest accuracy, indicating its robustness and reliability in accurately classifying the breast cancer classes based on the combined features. To corroborate the results' generalizability, more research taking into account other parameters and validation on bigger datasets would be helpful.

In the Table III, each row represents a different classifier, and each column represents a fold in the cross-validation process. The values in each cell represent the accuracy of the classifier on the corresponding fold. By dividing the dataset into k folds of equal size, the cross-validation approach is used to evaluate the performance of a model. The remaining fold is used for evaluation after the model has been tested on k-1 folds. Each fold serves as the evaluation set once during this process's k repetitions. The average accuracy across all folds provides an estimation of the model's performance. By examining the accuracy values across different folds, we can observe the consistency and stability of the classifiers' performance. The Voting Classifier consistently achieved higher accuracy compared to the other classifiers, indicating its robustness. Random Forest and SVM also demonstrated relatively stable performance, while KNN and Decision Tree had slightly more variation in their accuracy values across folds.

TABLE III. K-FOLD VALIDATION RESULTS

Classifier	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	Variance
Random Forest	0.86	0.88	0.90	0.85	0.87	0.872	0.0016
KNN	0.82	0.83	0.80	0.85	0.81	0.822	0.0007
SVM	0.85	0.87	0.84	0.86	0.88	0.858	0.0009
Decision Tree	0.75	0.77	0.73	0.78	0.76	0.758	0.0009
Voting Classifier	0.88	0.90	0.87	0.89	0.91	<b>0.89</b>	<b>0.0008</b>

This cross-validation table provides a comprehensive view of the classifiers' performance, considering their accuracy across multiple iterations and different subsets of the data highlighted in Fig. 4. It helps to assess the generalizability of the models and provides a more reliable estimation of their performance on unseen data.

## VI. CONCLUSION

Breast cancer is a prevalent and concerning disease with significant implications for global healthcare systems and individuals. Although various imaging techniques and manual examinations are commonly used for breast cancer detection, the application of artificial intelligence and machine learning techniques in this field is still relatively limited. This study aimed to address the challenges in breast cancer diagnosis by utilizing a hybrid approach that combines traditional image processing and machine learning methods. The proposed method incorporated feature fusion using GLCM, LBP, and histogram features, along with an ensemble learning approach

for improved classification performance. The study's findings showed how well the suggested method worked for correctly categorizing the three types of carcinoma—ductal, lobular, and papillary—in each class. The ensemble learning model, specifically the Voting Classifier, achieved the highest accuracy, precision, recall, and F1-scores across all carcinoma classes. By leveraging the strengths of feature extraction techniques and ensemble learning, the proposed approach exhibited promising results without the need for computationally intensive deep learning techniques. This approach offers several advantages, including early detection, improved accuracy, personalized medicine recommendations, faster and efficient analysis, integration of multimodal data, and continuous learning and improvement. Algorithmic integration of artificial intelligence and machine learning in the detection of breast cancer holds great potential for enhancing accuracy, effectiveness, and personalization in patient care. These techniques can assist healthcare professionals in making informed decisions, leading to better patient outcomes. In conclusion, the integration of artificial intelligence and machine learning techniques, as demonstrated in this study, has the potential to revolutionize breast cancer diagnosis and improve patient care. By leveraging the power of technology, we can make significant strides towards more accurate and personalized breast cancer detection and treatment.

## VII. FUTURE WORK

Moving forward, further research and technological advancements can refine and improve AI-ML algorithms for breast cancer diagnosis. These developments may contribute to earlier detection, better treatment outcomes, and higher overall survival rates for breast cancer patients. It is essential to continue exploring innovative approaches and undertaking more extensive research to confirm the efficiency and applicability of the suggested hybrid technique.

## REFERENCES

- [1] The global challenge of cancer, Nature Cancer, vol. 1, January 2020, pp. 1-2.
- [2] Arun Upadhyay, "Cancer: An unknown territory; rethinking before going ahead," Genes & Diseases, vol. 8, Iss. 5, September 2021, pp. 655-661.
- [3] He, Z., Chen, Z., Tan, M., Elingarami, S., Liu, Y., Li, T., Deng, Y., He, N., Li, S., Fu, J., and Li, W., "A review on methods for diagnosis of breast cancer cells and tissues," Cell Proliferation, vol. 53(7), July 2020: e12822.
- [4] Zahoor Saliha , Lali Ullah Ikram , Khan Attique Muhammad, Javed Kashif and Mehmood Waqar , "Breast Cancer Detection and Classification using Traditional Computer Vision Techniques: A Comprehensive Review", Current Medical Imaging, vol. 16(10), 2020, pp. 1187 – 1200.
- [5] Champaign JL and Cederbom GJ., "Advances in breast cancer detection with screening mammography," The Ochsner journal, vol. 2(1), January 2000 , pp. 33-5, PMID: 21765659.
- [6] Heang-Ping Chan, Ravi K. Samala and Lubomir M. Hadjiiski, "CAD and AI for breast cancer—recent development and challenges," The British Journal of Radiology, vol. 93(1108), 2020: 20190580
- [7] C. Kaushal, S. Bhat, D. Koundal and A. Singla, "Recent Trends in Computer Assisted Diagnosis (CAD) System for Breast Cancer Diagnosis Using Histopathological Images," IRBM, vol. 40, Iss. 4, August 2019, pp. 211-227.

- [8] Dileep G and Gianchandani Gyani S G, "Artificial Intelligence in Breast Cancer Screening and Diagnosis," *Cureus*, vol. 14(10), October 2022: e30318.
- [9] Marshall J, Peshkin B, Yoshino T et al., "The Essentials of Multiomics," *The Oncologist*, vol. 27, Iss. 4, April 2022, pp. 272–284.
- [10] Sebastian AM and Peter D., "Artificial Intelligence in Cancer Research: Trends, Challenges and Future Directions," *Life*, vol. 12(12), November 2022 :1991.
- [11] Hamamoto R, Suvarna K, Yamada M, Kobayashi K, Shinkai N, Miyake M, Takahashi M, Jinnai S, Shimoyama R, Sakai A, et al., "Application of Artificial Intelligence Technology in Oncology: Towards the Establishment of Precision Medicine," *Cancers*, vol. 12(12), November 2020:3532.
- [12] Zi-Hang Chen, Li Lin, Chen-Fei Wu, Chao-Feng Li, Rui-Hua Xu and Ying Sun, "Artificial intelligence for assisting cancer diagnosis and treatment in the era of precision medicine," *Cancer Communications*, vol. 41(11), November 2021, pp. 1100-1115.
- [13] Brindha Senthilkumar, Doris Zodinpuui, Lalawmpuii Pachuau, Saia Chenkual, John Zohmingthanga, Nachimuthu Senthil Kumar and Lal Hmingliana, "Ensemble Modelling for Early Breast Cancer Prediction from Diet and Lifestyle," *IFAC-PapersOnLine*, vol. 5(1), 2022, pp. 429-435.
- [14] Athraa H. Farhan and Mohammed Y. Kamil, "Texture Analysis of Breast Cancer via LBP, HOG, and GLCM techniques," *IOP Conference Series: Materials Science and Engineering*, vol. 928, 2nd International Scientific Conference of Al-Ayen University (ISCAU-2020) 15-16 July 2020, Thi-Qar, Iraq, 072098.
- [15] M. B. A. Rasyid, Yunidar, F. Arnia and K. Munadi, "Histogram statistics and GLCM features of breast thermograms for early cancer detection," 2018 International ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI-NCON), Chiang Rai, Thailand, 2018, pp. 120-124.
- [16] Amin Alqudah and Ali Mohammad Alqudah, "Sliding Window Based Support Vector Machine System for Classification of Breast Cancer Using Histopathological Microscopic Images," *IETE Journal of Research*, vol. 68:1, 2022, pp. 59-67.
- [17] Gour M, Jain S and Sunil Kumar T, "Residual learning based CNN for breast cancer histopathological image classification," *International Journal of Imaging Systems and Technology*, vol.30(3), September 2020, pp. 621-635.
- [18] Ziba Gandomkar, Patrick C. Brennan and Claudia Mello-Thoms, "MuDeRN: Multi-category classification of breast histopathological image using deep residual networks," *Artificial Intelligence in Medicine*, vol. 88, June 2018, pp. 14-24.
- [19] Beltran-Perez, C., Wei, HL. & Rubio-Solis, A., "Generalized Multiscale RBF Networks and the DCT for Breast Cancer Detection," *Int. J. Autom. Comput.*, vol.17, February 2020, pp. 55–70.
- [20] Abhinav Kumar, Sanjay Kumar Singh, Sonal Saxena, K. Lakshmanan, Arun Kumar Sangaiah, Himanshu Chauhan, Sameer Shrivastava and Raj Kumar Singh, "Deep feature learning for histopathological image classification of canine mammary tumors and human breast cancer," *Information Sciences*, vol. 508, January 2020, pp. 405-421.
- [21] Li, X., Shen, X., Zhou, Y., Wang, X., and Li, Q., "Classification of breast cancer histopathological images using interleaved DenseNet with SENet (IDSNet)," *PLOS ONE*, vol.15(5), May 2020: e0232127.
- [22] Duc My Vo, Ngoc-Quang Nguyen and Sang-Woong Lee, "Classification of breast cancer histology images using incremental boosting convolution networks," *Information Sciences*, vol. 482, May 2019, pp. 123-138.
- [23] Saxena S, Shukla S and Gyanchandani M, "Breast cancer histopathology image classification using kernelized weighted extreme learning machine," *International Journal of Imaging Systems and Technology*, vol. 31, no.1, March 2021, pp. 168-179.
- [24] Alom, M.Z., Yakopcic, C., Nasrin, M.S., Taha T. M. and Asari V. K., "Breast Cancer Classification from Histopathological Images with Inception Recurrent Residual Convolutional Neural Network," *J. Digit. Imaging*, vol. 32, August 2019, pp. 605–617.
- [25] Said Boumaraf, Xiabi Liu, Zhongshu Zheng, Xiaohong Ma and Chokri Ferkous, "A new transfer learning based approach to magnification dependent and independent classification of breast cancer in histopathological images," *Biomedical Signal Processing and Control*, vol. 63, January 2021: 102192.
- [26] Burçak, K.C., Baykan, Ö.K. and Uğuz, H., "A new deep convolutional neural network model for classifying breast cancer histopathological images and the hyperparameter optimisation of the proposed model," *J. Supercomput.*, vol. 77, January 2021, pp. 973–989.
- [27] Xie Juanying, Liu Ran, Luttrell Joseph and Zhang Chaoyang, "Deep Learning Based Analysis of Histopathological Images of Breast Cancer," *Frontiers in Genetics*, vol. 10, February 2019.
- [28] Jiang, Y., Chen, L., Zhang, H., and Xiao, X., "Breast cancer histopathological image classification using convolutional neural networks with small SE-ResNet module," *PLOS ONE*, vol. 14(3), March 2019:e0214587.
- [29] Zhongyi Han, Benzhenq Wei, Yuanjie Zheng, Yilong Yin, Kejian Li and Shuo Li, "Breast Cancer Multi-classification from Histopathological Images with Structured Deep Learning Model," *Scientific Reports*, vol. 7, June 2017:4172.
- [30] Rajesh Kumar, Rajeev Srivastava, and Subodh Srivastava, "Detection and Classification of Cancer from Microscopic Biopsy Images Using Clinically Significant and Biologically Interpretable Features," *Journal of Medical Engineering*, vol. 2015, August 2015: 457906.
- [31] Sheikh TS, Lee Y and Cho M., "Histopathological Classification of Breast Cancer Images Using a Multi-Scale Input and Multi-Feature Network," *Cancers*, vol. 12, no.8, July 2020:2031
- [32] Abdullah-Al Nahid, Mohamad Ali Mehrabi, and Yanan Kong, "Histopathological Breast Cancer Image Classification by Deep Neural Network Techniques Guided by Local Clustering," *BioMed Research International*, vol. 2018, March 2018:2362108.
- [33] Zhu, C., Song, F., Wang, Y. et al. Breast cancer histopathology image classification through assembling multiple compact CNNs. *BMC Med Inform Decis Mak*, vol. 19, 198, October 2019.
- [34] L. Gong, J. Yang and X. Zhang, "Semi-Supervised Breast Histological Image Classification by Node-Attention Graph Transfer Network," in *IEEE Access*, vol. 8, August 2020, pp. 158335-158345.
- [35] Kalpana George, Shameer Faziludeen, Praveen Sankaran and Paul Joseph K, "Breast cancer detection from biopsy images using nucleus guided transfer learning and belief based fusion," *Computers in Biology and Medicine*, vol. 124, September 2020:103954.
- [36] D.C.R. Novitasari, A. Lubab, A. Sawiji, A.H. Asyhar "Application of Feature Extraction for Breast Cancer using One Order Statistic, GLCM, GLRLM, and GLDM", *Advances in Science, Technology and Engineering Systems Journal*, vol. 4, no. 4, 2019, pp. 115-120.
- [37] Hao, Y., Zhang, L., Qiao, S., Bai, Y., Cheng, R., Xue, H., Hou, Y., Zhang, W., and Zhang, G., "Breast cancer histopathological images classification based on deep semantic features and gray level co-occurrence matrix," *PLOS ONE*, 17(5), May 2022 :e0267955.
- [38] S. J. A. Sarosa, F. Utaminigrum and F. A. Bachtiar, "Mammogram Breast Cancer Classification Using Gray-Level Co-Occurrence Matrix and Support Vector Machine," 2018 International Conference on Sustainable Information Engineering and Technology (SIET), Malang, Indonesia, 2018, pp. 54-59.
- [39] R. Touahri, N. AzizI, N. E. Hammami, M. Aldwairi and F. Benaïda, "Automated Breast Tumor Diagnosis Using Local Binary Patterns (LBP) Based on Deep Learning Classification," 2019 International Conference on Computer and Information Sciences (ICCIS), Sakaka, Saudi Arabia, 2019, pp. 1-5.
- [40] Ahirrao, Sonal R. and Bormane, D. S., "A novel approach for Face Recognition using Local Binary Pattern," *International Journal of Image Processing and Vision Science*, vol. 1 : Iss. 1 , Article 6, July 2012.
- [41] Ü. Budak and A.B. Güzel, "Automatic Grading System for Diagnosis of Breast Cancer Exploiting Co-occurrence Shearlet Transform and Histogram Features," *IRBM*, vol. 41(2), April 2020, pp. 106-114.
- [42] F. A. Spanhol, L. S. Oliveira, C. Petitjean and L. Heutte, "A Dataset for Breast Cancer Histopathological Image Classification," in *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 7, July 2016, pp. 1455-1462.
- [43] M. Adel, A. Kotb, O. Farag, M. S. Darweesh and H. Mostafa, "Breast Cancer Diagnosis Using Image Processing and Machine Learning for

- Elastography Images," 2019 8th International Conference on Modern Circuits and Systems Technologies (MOCASST), Thessaloniki, Greece, 2019, pp. 1-4.
- [44] S. H. Kassani, P. H. Kassani, M. J. Wesolowski, K. A. Schneider and R. Deters, "Breast Cancer Diagnosis with Transfer Learning and Global Pooling," 2019 International Conference on Information and Communication Technology Convergence (ICTC), Jeju, Korea (South), 2019, pp. 519-524.
- [45] S. Punitha, A. Amuthan and K. Suresh Joseph, "Benign and malignant breast cancer segmentation using optimized region growing technique," Future Computing and Informatics Journal, vol. 3(2), December 2018, pp. 348-358.
- [46] D. A. Zebari, H. Haron, D. M. Sulaiman, Y. Yusoff and M. N. Mohd Othman, "CNN-based Deep Transfer Learning Approach for Detecting Breast Cancer in Mammogram Images," 2022 IEEE 10th Conference on Systems, Process & Control (ICSPC), Malacca, Malaysia, 2022, pp. 256-261
- [47] H. -C. Lu, E. -W. Loh and S. -C. Huang, "The Classification of Mammogram Using Convolutional Neural Network with Specific Image Preprocessing for Breast Cancer Detection," 2019 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD), Chengdu, China, 2019, pp. 9-12
- [48] Cengiz, Enes, Kelek, Muhammed Mustafa, Oğuz, Yüksel and Yılmaz, Cemal, "Classification of breast cancer with deep learning from noisy images using wavelet transform," Biomedical Engineering / Biomedizinische Technik, vol. 67, no. 2, March 2022, pp. 143-150.
- [49] Murcia-Gómez D, Rojas-Valenzuela I and Valenzuela O, "Impact of Image Preprocessing Methods and Deep Learning Models for Classifying Histopathological Breast Cancer Images," Applied Sciences, vol. 12, no. 22, November 2022:11375.
- [50] Kode H and Barkana BD, "Deep Learning- and Expert Knowledge-Based Feature Extraction and Performance Evaluation in Breast Histopathology Images," Cancers, vol. 15, no.12, June 2023: 3075.
- [51] Assiri AS, Nazir S and Velastin SA, "Breast Tumor Classification Using an Ensemble Machine Learning Method," Journal of Imaging, vol.6, no.6, May 2020:39.
- [52] Aswathy, M.A. and Jagannath, M, "An SVM approach towards breast cancer classification from H&E-stained histopathology images based on integrated features," Med Biol Eng Comput, vol. 59, July 2021, pp. 1773–1783.
- [53] M. Samridha Majumdar, Payel Pramanik and Ram Sarkar, "Gamma function based ensemble of CNN models for breast cancer detection in histopathology images," Expert Systems with Applications, vol. 213, Part B, March 2023:119022.

# Deep Conv-LSTM Network for Arrhythmia Detection using ECG Data

Alisher Mukhametkaly<sup>1</sup>, Zeinel Momynkulov<sup>2</sup>, Nurgul Kurmanbekkyzy<sup>3</sup>, Batyrkhan Omarov<sup>4</sup>

International Information Technology University, Almaty, Kazakhstan<sup>1, 2, 4</sup>

Kazakh-Russian Medical University, Almaty, Kazakhstan<sup>3</sup>

Al-Farabi Kazakh National University, Almaty, Kazakhstan<sup>4</sup>

NARXOZ University, Almaty, Kazakhstan<sup>4</sup>

INTI International University, Kuala Lumpur, Malaysia<sup>4</sup>

**Abstract**—In the evolving realm of medical diagnostics, electrocardiogram (ECG) data stands as a cornerstone for cardiac health assessment. This research introduces a novel approach, leveraging the capabilities of a Deep Convolutional Long Short-Term Memory (Conv-LSTM) network for the early and accurate detection of arrhythmias using ECG data. Traditionally, cardiac anomalies have been diagnosed through heuristic means, often requiring intricate scrutiny and expertise. However, the Deep Conv-LSTM model proposed herein addresses the inherent limitations of traditional methods by amalgamating the spatial feature extraction capability of convolutional neural networks (CNN) with the temporal sequence learning capacity of LSTM networks. Initial results derived from a diverse dataset, comprising myriad ECG waveform anomalies, delineated an enhancement in accuracy, reducing false positives and facilitating timely interventions. Notably, the model showcased adaptability in handling the burstiness of ECG signals, reflecting various heart rhythms, and the perplexity inherent in diagnosing subtle arrhythmic events. Additionally, the model's ability to discern longer, more complex patterns alongside transient anomalies offers potential for broader applications in telemetry and continuous patient monitoring systems. It is anticipated that this innovative fusion of CNN and LSTM architectures will usher a paradigm shift in automated arrhythmia detection, bridging the chasm between technology and the intricate nuances of cardiac physiology, thus improving patient outcomes.

**Keywords**—Deep learning; Conv-LSTM; classification; ECG; CNN

## I. INTRODUCTION

The landscape of medical diagnostics has been transformed by technological advancements, and at the heart of this transformation lies the persistent quest to enhance the accuracy, efficiency, and predictability of diagnostic tools [1]. One of the most pivotal diagnostic tools in cardiology is the electrocardiogram (ECG), a non-invasive method capturing the electrical activity of the heart over a specified period. ECG data, with its intricate waveforms, provides clinicians with invaluable insights into the rhythmic and conduction anomalies of the heart [2]. However, the challenge has always been the interpretation of this data, particularly in recognizing subtle or transient arrhythmic events, which often elude detection or result in misdiagnoses.

Historically, the primary approach to ECG interpretation has been manual, relying heavily on the expertise and acumen of medical professionals. While this heuristic methodology has served for decades, it is not devoid of limitations. The manual interpretation is not only time-consuming but is also vulnerable to human error, particularly when confronted with vast volumes of continuous monitoring data or nuanced arrhythmic events that may get obscured amidst the background noise [3]. Furthermore, in scenarios where immediate interventions are crucial, delays in detection can potentially compromise patient outcomes.

In the wake of these challenges, the fusion of computational methods and medical diagnostics has emerged as a promising frontier [4]. The last two decades have witnessed a surge in the adoption of machine learning techniques for medical data interpretation, specifically in cardiology. Among these, neural networks, due to their innate ability to learn complex patterns, have shown promise in ECG data interpretation [5]. But with the rich temporal structure of ECG signals, a mere feed-forward neural network might not suffice. Enter the realm of recurrent neural networks (RNN), with their ability to learn sequences [6], and their more sophisticated counterpart, the Long Short-Term Memory (LSTM) networks [7]. LSTMs, with their intricate architecture, have the capacity to remember and learn from long-term dependencies in data, making them apt for ECG waveform analysis.

However, the story doesn't end there. ECG data, with its nuances, presents both spatial and temporal challenges. While LSTMs aptly address the temporal aspects, spatial feature extraction becomes a stumbling block [8]. This is where convolutional neural networks (CNN) come into the picture. Renowned for their prowess in spatial feature extraction, especially in image data, CNNs can discern patterns in localized data regions. The logical progression, then, was the integration of these two potent architectures, leading to the advent of Convolutional LSTM (Conv-LSTM) networks [9]. By harnessing the spatial feature extraction capabilities of CNNs and the temporal pattern learning of LSTMs, Conv-LSTMs offer a balanced approach to sequence data with spatial intricacies, like ECG waveforms.

The present research introduces a Deep Conv-LSTM model tailored for the detection of arrhythmias from ECG data. The



ambition driving this study is twofold: first, to address the aforementioned challenges in ECG interpretation by reducing false positives and false negatives; and second, to offer a robust, scalable, and efficient model that can be seamlessly integrated into real-time patient monitoring systems, thus paving the way for timely and effective clinical interventions.

In subsequent sections, this paper will delve deep into the architecture of the proposed model, detailing its layers, parameters, and training regimen. A comprehensive evaluation, juxtaposing the model against traditional methods and other machine learning approaches, will underscore its effectiveness. Furthermore, a detailed discussion on its potential applications, adaptability, and future directions will round off this exploration into the potential of Deep Conv-LSTM networks in revolutionizing arrhythmia detection using ECG data.

The convergence of cardiology and computational methods, as seen through the lens of this research, heralds an era where technology doesn't just aid, but actively augments and refines, the capabilities of clinicians, promising improved diagnostic accuracy and better patient outcomes.

## II. RELATED WORKS

The confluence of cardiology and computational methods, especially in the realm of electrocardiogram (ECG) data interpretation, has seen considerable research attention in recent decades. As our understanding of ECG data's depth and complexity has grown, so too has the need for accurate, efficient, and scalable analysis methods. This section reviews related works that have hitherto shaped the landscape of automated ECG interpretation, particularly focusing on machine learning techniques and their application to arrhythmia detection.

### A. Traditional ECG Interpretation Techniques

Before the widespread adoption of computational models, traditional ECG interpretation largely leaned on signal processing techniques. Pan and Tompkins (1985) proposed an algorithm based on derivative, integration, and thresholding methods for QRS detection [10]. Though seminal and widely adopted, its deterministic nature limited its ability to adapt to diverse ECG morphologies.

The foundation of traditional ECG interpretation revolves around the identification and examination of waveform components: the P wave, QRS complex, and T wave. By analyzing the amplitude, duration, and morphological attributes of these components, clinicians could infer various cardiac functionalities, such as atrial and ventricular depolarization and repolarization.

Given that the QRS complex is the most prominent feature on an ECG tracing, much of the early research in automated ECG interpretation honed in on its accurate detection. The Pan and Tompkins algorithm, as previously mentioned, became a seminal work in this space. Their method combined bandpass filtering, differentiation, squaring, and integration to emphasize the QRS complex's characteristics and subsequently detect it using a threshold mechanism. This approach achieved remarkable accuracy for its time and laid the groundwork for many succeeding algorithms.

### B. Neural Networks in ECG Analysis

With the rise of artificial neural networks (ANN), attempts were made to employ them for ECG interpretation. Acharya et al. (2017) provided a comprehensive survey on the use of ANN in detecting cardiac disorders [11]. While ANN models demonstrated promise, they lacked the ability to exploit the temporal dependencies intrinsic to ECG data.

For effective ANN-based ECG analysis, the extraction of salient features from raw ECG data was paramount. Techniques such as wavelet transform, Fourier transform, and principal component analysis (PCA) were employed to distill relevant information from the ECG waveform, which was then fed into the neural networks for classification [12].

While ANNs exhibited potential, their early applications in ECG analysis faced challenges. Overfitting, where the model performed exceptionally well on training data but poorly on unseen data, was a recurrent issue [13]. Moreover, the lack of interpretability of ANNs posed challenges in clinical adoption, as physicians often sought explanations for diagnostic decisions.

Addressing the limitations of early ANN applications, researchers introduced regularization techniques like dropout and early stopping to combat overfitting. Furthermore, optimization strategies, such as adaptive learning rates and momentum, were employed to hasten and improve the training process.

### C. Advent of Recurrent Neural Networks (RNN)

Understanding the temporal nature of ECG signals, researchers began to explore RNNs. This architecture facilitates the retention of previous data points in the sequence, rendering RNNs uniquely apt for tasks necessitating memory of past inputs, such as time-series analysis, language modeling, and, notably, ECG signal processing [14]. However, the traditional RNNs faced challenges in learning long-term dependencies due to issues like vanishing gradient, leading to the exploration of more sophisticated architectures.

In the context of ECG, RNNs began showing promise in detecting cardiac anomalies that are heavily reliant on temporal patterns. For instance, atrial fibrillation, a disorder characterized by rapid and irregular heartbeats, could be better identified when considering the preceding cardiac activity [15]. RNNs were proficient in capturing these long-term dependencies and variations in heart rhythms.

### D. LSTM Networks for ECG Interpretation

LSTMs, designed to overcome the shortcomings of traditional RNNs, quickly became the architecture of choice for sequence data like ECG. Xie et al. (2018) employed LSTMs for atrial fibrillation detection from short single-lead ECG records, demonstrating a marked improvement in accuracy over traditional algorithms [16].

To address the limitations of vanilla RNNs, Hochreiter and Schmidhuber introduced Long Short-Term Memory (LSTM) networks in 1997 [17]. LSTM units are equipped with gates that regulate the flow of information, making them adept at learning and remembering over long sequences, thus addressing the shortcomings of standard RNNs. In ECG

analysis, LSTM's capability to capture long-term dependencies improved the accuracy and robustness of rhythm classifications and anomaly detections.

#### E. Convolutional Neural Networks (CNN)

Parallely, CNNs gained traction, especially for spatial feature extraction. Kiranyaz et al. (2016) employed 1-D CNNs for ECG classification, harnessing the architecture's ability to discern localized patterns [18]. While CNNs adeptly tackled spatial complexities, they were less suited for the intricate temporal patterns in ECG data.

Given the need to extract local features in ECG signals before identifying temporal patterns, a hybrid architecture merging Convolutional Neural Networks (CNN) with LSTMs began gaining traction. CNNs excel at local pattern recognition, identifying intricate waveform shapes in ECG data. Their integration with LSTMs resulted in models that could process ECG signals with remarkable precision, capturing both spatial and temporal dependencies.

#### F. Hybrid Models - Combining CNNs and RNNs

Given the strengths and limitations of CNNs and RNNs, researchers embarked on efforts to combine the two. Rajpurkar et al. (2017) presented a model that used a combination of CNNs and a gated recurrent unit (GRU) to detect multiple arrhythmia types from ECG data [19]. Their work underscored the potential of hybrid models, setting the stage for further exploration.

The rapid expansion of deep learning in the field of ECG analysis led researchers to experiment with combining the strengths of different neural architectures. One such fusion is that of Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), which synergistically harness the spatial feature extraction capability of CNNs with the temporal pattern recognition of RNNs [20]. This hybrid approach quickly emerged as a powerful tool for decoding ECG signals with unparalleled precision.

Typically, a hybrid model begins with one or more convolutional layers to process raw ECG signals. These layers effectively identify key patterns and anomalies within beats. The extracted features are then passed to RNN or LSTM layers, which analyze the interdependencies between these features and ascertain longer-term anomalies or rhythms present in the sequence.

Hybrid models have consistently demonstrated superior performance in ECG classification tasks over using CNNs or RNNs independently. By addressing both intra-beat and inter-beat dependencies, they can detect a wider range of cardiac anomalies with greater accuracy.

#### G. Conv-LSTM in Biomedical Signal Processing

Convolutional Long Short-Term Memory (Conv-LSTM) networks emerged as an innovative deep learning architecture, adept at handling spatiotemporal data [21]. Rooted in the hybridization of CNNs and LSTMs, Conv-LSTM extends the concept to fuse convolutional operations directly into the recurrent gates of LSTM. This has profound implications for biomedical signal processing, particularly in ECG analysis, given the intricate interplay of spatial and temporal data.

The Conv-LSTM, introduced by Xingjian Shi et al. in 2015, modifies traditional LSTM units by replacing the matrix multiplications with convolutional operations [20]. This ensures that both spatial (localized features within data) and temporal dependencies (order and sequence of data) are concurrently processed, a trait indispensable for biomedical signals.

ECG signals represent a series of heartbeats over time. The shape of individual beats (P, Q, R, S, T waves) encodes spatial information, while the order and rhythm of these beats capture temporal information [21]. Conv-LSTM, with its innate capacity to process both, offers a robust framework for ECG signal analysis.

#### H. Challenges and Opportunities

Despite advancements, several challenges persist in automated ECG analysis. Noise, artifacts, and inter-patient variability often confound even sophisticated models [22]. Moreover, the need for vast labeled datasets for training remains a bottleneck. Transfer learning, domain adaptation, and unsupervised learning present exciting frontiers, potentially reducing the need for vast labeled datasets [23].

While deep learning methods showcases significant promise, there exist challenges in optimizing network parameters and ensuring computational efficiency. Yet, with advances in GPU technologies and refined training techniques, it's anticipated that Conv-LSTM will cement itself as a cornerstone in biomedical signal processing [24].

#### I. Real-world Applications

Beyond pure academic exploration, there is a growing body of work dedicated to integrating these advanced models into real-world systems. Wearable health tech, telemedicine platforms, and continuous monitoring systems in clinical settings are actively exploring the integration of models like Conv-LSTMs [25].

In light of the above, our research into the Deep Conv-LSTM Network for arrhythmia detection is positioned at the intersection of past learnings and future potential. Recognizing the strengths and limitations of prior works, our endeavor is to present a model that not only showcases superior performance metrics but also addresses some of the persistent challenges in the realm of automated ECG analysis.

### III. MATERIALS AND METHODS

Advanced therapeutic strategies are paramount in enhancing therapeutic results for cardiovascular ailment patients. Conventional therapeutic modalities typically bifurcate into two categories: hands-on therapeutic intervention and robotics-facilitated methodologies [26]. These prevailing techniques, however, grapple with distinct challenges. Notwithstanding their sophisticated functionality, robotics solutions come with elevated acquisition and upkeep expenditures, thus challenging their wide-scale adoption. On the other hand, the efficiency of both synthetic and hands-on therapeutic regimens is often hampered by the continuous dearth of healthcare practitioners.

Additionally, cardiovascular ailment-oriented rehabilitation is characteristically an extended endeavor. The protracted nature of this process, when juxtaposed with the inherent obstacles of the existing paradigms, accentuates the imperative for a more sustainable methodology. Such a method should ideally be insulated from exorbitant technological investments or overextended medical personnel yet should be adept at furnishing indispensable rehabilitative care to the patients [27].

In response to this evident lacuna, a contemporary paradigm has surfaced: an autonomous rehabilitative training framework. This blueprint, meticulously crafted, pivots around the contemporary Human Activity Recognition (HAR) paradigms, with the intent to discern and mentor patients throughout their therapeutic routines [28]. By amalgamating

automation's tenets with rehabilitative principles, this framework is on the cusp of transforming the cardiovascular ailment therapeutic landscape, endorsing patient recuperation in a more democratized and economically prudent fashion.

A visual representation of the suggested algorithmic structure is delineated in Fig. 1, elucidating its potential in automating cardiovascular ailment therapeutic regimens. Through its proficiency in identifying and supervising therapeutic activities, the algorithm proffers real-time counsel and oversight to patients, fortifying the correctness and regularity of their exercises. Such an initiative has the latent capacity to bolster the efficacy and reach of cardiovascular ailment therapy, rendering it an apt countermeasure for this pressing healthcare predicament.

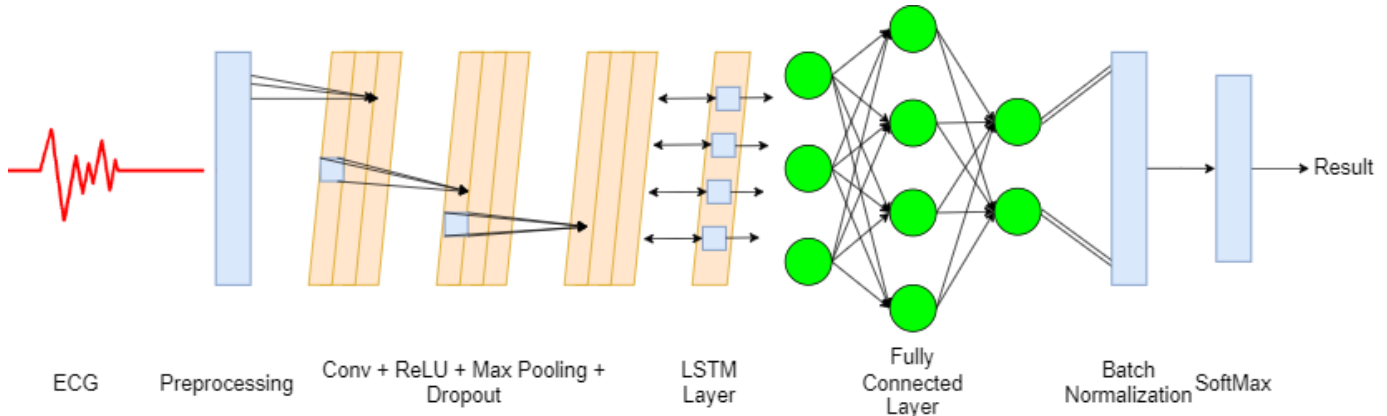


Fig. 1. The proposed Conv-LSTM Network for arrhythmia detection.

The advanced ECG Conv-LSTM framework endeavors to capitalize on the synergistic merits of integrating convolutional neural networks with LSTM. This overarching ambition bifurcates into dual facets: discerning electrocardiograms and subsequently classifying them. The inception phase of this inquiry is committed to data curation, dimensionality curtailment, and preliminary processing. Subsequently, we delve into the attributes of electrocardiograms, employing a medley of profound learning modalities to augment their classification potency. Multiple trials centered on ECG recognition and classification were undertaken to assess the efficacy of the proposed model. The quintessential elements of the algorithm will be elucidated and critically appraised in ensuing segments.

#### A. Convolutional Neural Network

The advanced ECG Conv-LSTM framework falls under the domain of neural architectures termed deep neural networks, characterized by their multilayered composition [29]. This model drew inspiration from the synergistic blend of the receptive field and computational acumen, exhibiting greater intricacy compared to traditional neural constructs. Models rooted in the deep neural network paradigm, endowed with supplementary layers, can attain a depth of learning surpassing that of their simpler counterparts.

Convolutional neural networks (CNNs), owing to their spatial structuring and weight allocation strategy, exhibit a commendable resilience against deformations [30], rendering them apt for tasks associated with image analysis. The

principle of weight-sharing inherent to CNNs not only simplifies the model architecture but also augments operational efficacy and astutely calibrates the weight count. Accepting image datasets, CNNs scrutinize them, subsequently projecting precise categorizations of the image type predicated on the evaluated data. These input visuals are epitomized as bidimensional vectors, a format adeptly managed by CNNs.

Within the outlined ECG Conv-LSTM architecture, the CNN component is instrumental in distilling salient features. This research employs LSTM to segment the fed ECG data into distinct clusters. A subsequent segment furnishes an exhaustive elucidation of the convolutional neural network's role in feature distillation. The CNN's training regimen is encapsulated algebraically in Eq. (1), wherein  $Z_i$  signifies the input collection,  $W_i$  denotes the weight assortment, and  $B$  symbolizes the bias mechanism.

$$P = f\left(\sum_{i=1}^N Z_i \cdot W_i + B\right) \quad (1)$$

#### B. Long Short-Term Memory Network

Within the sophisticated ECG Conv-LSTM architecture, the role of LSTM is pivotal in circumventing complications such as gradient diminution or exacerbation throughout the training phase. To regulate the weights, the technique of backpropagation (BP) is instituted. This method inaugurates by deducing the gradient via the chain principle. Subsequent to this, a systematic recalibration of weights ensues, based on the

discerned loss. The inception of backpropagation is at the neural network's output stratum, and as weight updating transpires, it cascades towards the initial layer, potentially giving rise to issues like the attenuating or inflating gradients [31].

Proposing a remedy to the aforementioned gradient diminution challenge, inherent in standard recurrent neural networks, the LSTM strategy comes to the forefront. Distinct from conventional recurrent neural structures, LSTM possesses an adeptness in retaining extended data sequences efficaciously. Essentially, LSTM embodies a recurrent neural architecture but is augmented with supplementary memory modules, empowering it to encapsulate and preserve pivotal data across elongated sequences [32].

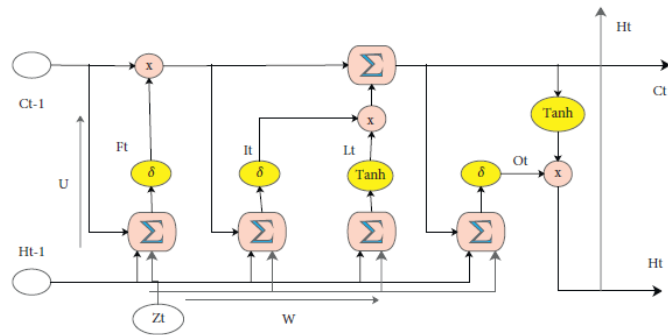


Fig. 2. LSTM block.

Illustrated in Fig. 2 is the architecture of LSTM, tailored to assimilate and perpetuate knowledge from sequences spanning extended durations. The LSTM framework is delineated into four cardinal components: the input gate ( $I_t$ ), the output gate ( $O_t$ ), the forget gate ( $F_t$ ), and the cell state ( $C_t$ ) pegged to a particular temporal juncture ( $t$ ) [33]. The state vector,  $C_{t-1}$ , enshrines information from the antecedent phase. Predicated upon the freshest influx of data, determinations regarding weight modifications are reached. The vector  $L_t$  articulates the data stemming from the preceding input. The time- $t$  specific input vector is represented as  $Z_t$ . The output emanating from the pertinent cells is encapsulated in  $H_t$  and  $H_{t-1}$ , while the memory cells are symbolized by  $C_t$  and  $C_{t-1}$ . The weight attributes of the quartet of gates— $I_t$ ,  $O_t$ ,  $F_t$ , and  $C_t$ —are mirrored in  $W$  and  $U$ . Owing to its intrinsic design, LSTM is poised to adeptly decode intricate data sequences.

$$L_t = \tanh(Z_t \cdot W_L + H_{t-1} \cdot U_L) \quad (2)$$

$$F_t = \sigma(Z_t \cdot W_F + H_{t-1} \cdot U_F) \quad (3)$$

$$I_t = \sigma(Z_t \cdot W_I + H_{t-1} \cdot U_I) \quad (4)$$

$$O_t = \sigma(Z_t \cdot W_O + H_{t-1} \cdot U_O) \quad (5)$$

$$C_t = F_t \cdot C_{t-1} + I_t \cdot L_t \quad (6)$$

$$H_t = O_t \cdot \tanh(C_t) \quad (7)$$

The symbols  $\sigma$  and  $\tanh$  denote nonlinear activation functions, while the weight parameters  $U_I, W_I, U_F, W_F, U_O, W_O, U_L, W_L$  each exhibit dimensions of  $M \times 2N$ . Here,  $M$  epitomizes the count of memory cells, and  $N$  delineates the dimensionality of the input vector. A comprehensive elucidation of the LSTM's mathematical underpinnings, pivotal to its operational framework, is documented in [34], specifically referenced in Eq. (2) to (7).

#### IV. EXPERIMENTAL SETUP AND RESULTS

In the subsequent section, we present the outcomes of our empirical investigations. These results have been meticulously extracted and analyzed to shed light on the efficacy and nuances of the proposed model. Beyond mere data, they provide invaluable insights into the performance, challenges, and potential optimizations for the methods under scrutiny. As we delve into this segment, readers are invited to evaluate the results in the broader context of our research objectives and the prevailing literature in the field.

##### A. Data

For the assessment of the advanced model put forth, we leveraged the ECG arrhythmia classification repository [35]. The ECG Arrhythmia Classification Repository stands as an exhaustive compilation that delves deep into the myriad nuances of cardiac irregularities. It encapsulates twelve primary cardiac rhythm categories, encompassing, but not restricted to, sinus rhythm, atrial fibrillation, and ventricular escape rhythm. This compilation offers a formidable platform for scrutinizing a gamut of cardiac aberrations, especially accentuating intricate and often obscured states such as ventricular fibrillation.

Beyond the elemental ECG traces, the repository furnishes an array of derived attributes. This gamut spans metrics like heart rate fluctuations, attributes of the Q, R, and S complexes, and variations in the T-wave, augmenting the repository's diagnostic potential. Despite its voluminous nature and array of diverse metrics, the repository has been curated with meticulous precision, ensuring ease of navigation and utility. Encompassing myriad ECG traces from a vast demographic spectrum further accentuates the data's depth and adaptability.

The ECG Arrhythmia Classification Repository underscores its pivotal role in propelling insights into cardiovascular health. Serving as an indispensable instrument for academicians and clinicians, it facilitates the exploration and comprehension of a plethora of arrhythmic manifestations. Furthermore, it paves the way for precocious detection, refined diagnostic procedures, and optimized cardiac therapeutic interventions, thus heralding prospective strides in cardiological research and application. Fig. 3 demonstrates data samples of electrocardiograms that used in this study.

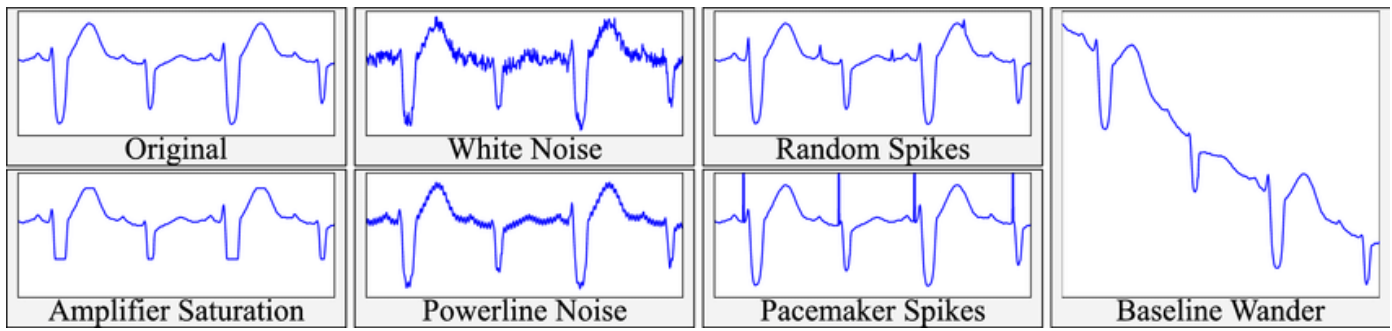


Fig. 3. Data samples.

### B. Evaluation Parameters

In the ensuing subsection, we direct our focus towards the evaluation parameters, the bedrock upon which our research findings stand. These parameters, carefully chosen and calibrated, are instrumental in gauging the effectiveness, precision, and reliability of our proposed model. By shedding light on these metrics, we aim to provide readers with a clear understanding of the benchmarks against which our results are measured, and the criteria that underpin our analyses. It is crucial to grasp the intricacies of these parameters to fully comprehend the depth and significance of the results presented. Let us delve deeper into the specifics of these evaluation metrics and their pivotal role in shaping our research narrative.

In the realm of classification tasks, particularly in scenarios where consequences of misclassification can be severe, precision emerges as a paramount metric. Precision, often referred to as the positive predictive value, is a measure of a model's accuracy in terms of its positive predictions. In simpler terms, it answers the question: Of all the instances that the model predicted as positive, how many were genuinely positive? Mathematically, precision can be articulated as in Eq. (8) [36]:

$$precision = \frac{TP}{TP + FP} \quad (8)$$

True Positives (TP): The count of positive instances correctly predicted as positive by the model.

False Positives (FP): The count of negative instances incorrectly predicted as positive.

In the intricate tapestry of classification metrics, Recall—often synonymous with Sensitivity or True Positive Rate—holds a pivotal position. As an evaluative criterion, Recall is centered around the model's adeptness in identifying all relevant instances within the dataset. Formally, Recall is described as in Eq. (9) [37]:

$$recall = \frac{TP}{TP + FN} \quad (9)$$

In essence, Recall quantifies the proportion of actual positives that were accurately captured by the model. High recall indicates that the classifier successfully identified most of the positive cases, minimizing the chances of type II errors or false negatives.

The salience of Recall becomes particularly pronounced in scenarios where overlooking positive instances bears substantial consequences. To illustrate, in the domain of medical diagnosis, missing a true case of a disease (resulting in a false negative) can have grave repercussions, from delayed treatment to reduced patient survival rates. In such contexts, achieving elevated levels of Recall becomes paramount, even if it sometimes comes at the expense of Precision.

Amid the pantheon of evaluative metrics in classification, the F-score, often termed the F1 score, emerges as a harmonized measure that synthesizes both Precision and Recall into a singular, cohesive metric. As such, it provides a more holistic representation of a model's performance, especially in scenarios where an equal weight is ascribed to both false positives and false negatives. Mathematically, the F-score is defined as in Eq. (10) [38]:

$$Fscore = \frac{2 \cdot precision \cdot recall}{precision + recall} \quad (10)$$

This formulation effectively captures the harmonic mean of Precision and Recall. Unlike the arithmetic mean, the harmonic mean gives a more conservative estimate and tends towards the smaller of the two values. Thus, a model can only achieve a high F-score if both Precision and Recall are high, ensuring a balanced performance [39].

The F-score's significance is particularly accentuated in situations with imbalanced datasets, where one class may heavily outnumber the other. In such contexts, sheer accuracy can be misleading, as a model might achieve high accuracy by merely predicting the majority class. The F-score, by virtue of its dependence on both Precision and Recall, provides a more nuanced and rigorous assessment of the model's capabilities.

While the F1 score gives equal weight to Precision and Recall, the broader family of F-scores allows for differential weighting [40]. The generalized F $\beta$ -score is given by Eq. (11):

$$F\beta - score = \left(1 + \beta^2\right) \frac{precision \cdot recall}{\left(\beta^2 \times precision\right) + recall} \quad (11)$$

Where  $\beta$  determines the relative weight given to Precision compared to Recall. A  $\beta$  value greater than 1 prioritizes Recall, while a value less than 1 accentuates Precision.

In conclusion, the F-score serves as an indispensable metric, elegantly amalgamating the distinct yet intertwined

dimensions of Precision and Recall. It proffers a comprehensive, balanced view of model performance, making it a vital tool in the evaluative arsenal of machine learning and data analytics endeavors.

C. Results

Navigating into the crux of our investigation, this subsection unveils the empirical findings derived from our meticulously crafted experiments. Grounded in rigorous methodologies and analytical rigor, the results illuminate the performance and efficacy of the proposed model vis-à-vis the outlined objectives. By dissecting these outcomes, we endeavor to provide a lucid understanding of the model's capabilities,

shedding light on its strengths and potential areas of refinement. Readers are invited to traverse this analytical journey, parsing the data and insights presented, to glean a comprehensive understanding of the model's real-world applicability and significance.

Fig. 4 delineates the confusion matrices, juxtaposing each category against the baseline "normal" class. Evidently, the class denoted as "hypertension" emerges with superior classification precision relative to its counterparts. Broadly, the categorization across classes manifests commendable accuracy in the realm of cardiac disease classification.

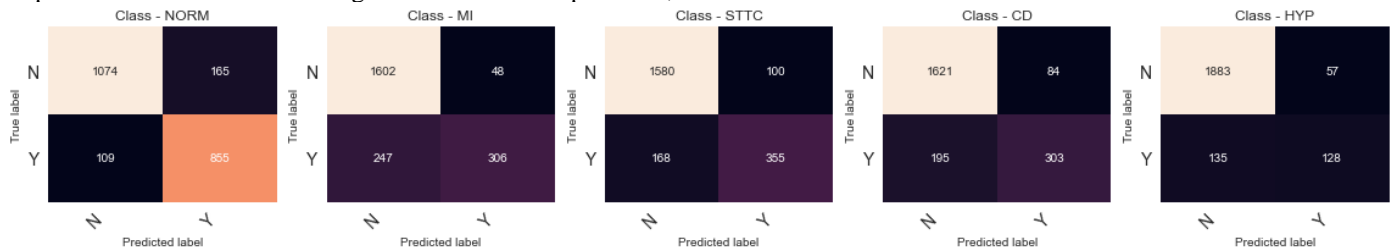


Fig. 4. Confusion matrix.

Fig. 5 presents the performance metrics of the proposed Conv-LSTM architecture across 40 learning iterations. The blue trajectory delineates the accuracy attained during the training phase, while the orange trajectory captures the testing phase's accuracy as a function of training iterations. Upon completing 40 iterations, the model registered a training accuracy of 88% and a testing accuracy of 86%. Furthermore, the insights suggest that reaching an optimal classification accuracy for cardiac conditions can be achieved within 20 learning epochs.

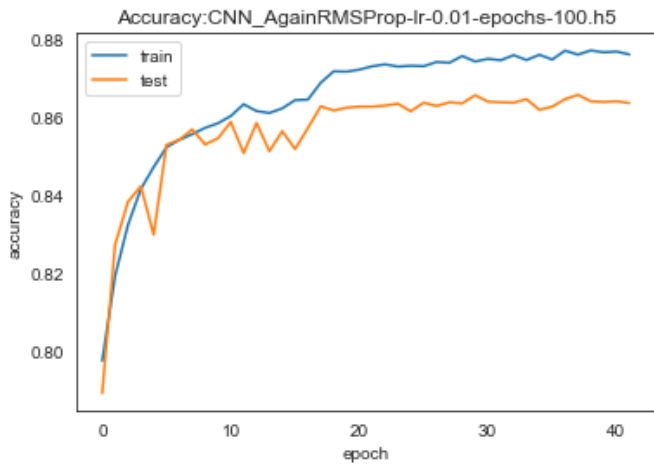


Fig. 5. Training and validation accuracy.

In a parallel context, Fig. 6 provides insights into both training and validation losses over 40 learning iterations. The findings delineate an inversely proportional relationship between accuracy metrics and the respective loss values. As the number of epochs amplifies, there's a discernible decrement in both training and validation losses. Echoing prior observations, optimizing the model's performance—both in

terms of maximal accuracy and minimal loss—appears achievable within a span of 20 epochs.

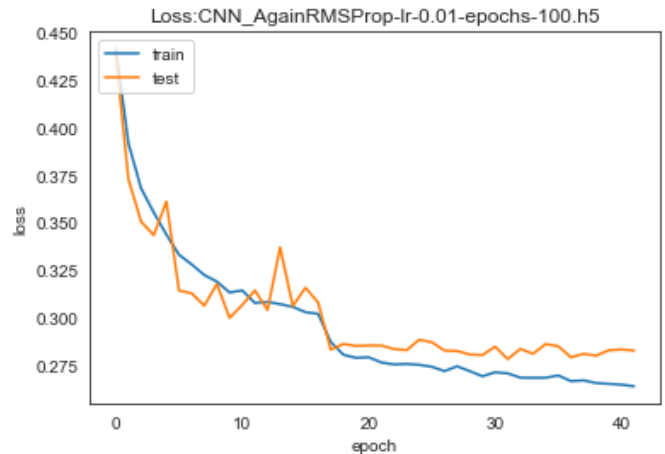


Fig. 6. Training and validation loss.

The efficacy of our proposed 3D deep Conv-LSTM network was critically assessed for its proficiency in heart disease detection using ECG datasets. Drawing direct comparisons between our results and prior studies demands caution, given the variability in test set sizes and specific heart disease types addressed. Nonetheless, our innovative methodology surpassed many existing benchmarks in terms of accuracy, marking a forward leap in heart disease classification.

V. DISCUSSION

The current exploration into the realm of ECG data analysis reveals evolving trends in both data acquisition techniques and algorithmic modeling. By investigating the performance of the 3D deep Conv-LSTM network in this context, a broader understanding emerges of the potential avenues for ECG-based

heart disease detection and the future trajectory of this research domain.

#### A. Emerging Trends

Cardiac health, especially the early detection of potential problems using ECG data, has become an area of increasing interest for researchers and clinicians alike. Several trends underpin this:

**Granular Data Collection:** With advancements in wearable technology and remote monitoring, there has been a surge in the volume of ECG data available for analysis [41]. This has catalyzed a move towards more complex algorithms capable of handling vast datasets and extracting meaningful patterns.

**Interdisciplinary Collaborations:** The melding of expertise from the realms of cardiology, biomedical engineering, and machine learning has fueled innovation, with each domain offering unique insights that enrich the overall analytical process [42].

**Real-time Monitoring:** As healthcare pivots towards a more preventive approach, there's an increased focus on real-time monitoring and instantaneous analysis [43]. This has spurred a shift from conventional post-test evaluations to immediate, actionable insights from ECG data.

#### B. Generalization of Results

The versatility of the proposed 3D deep Conv-LSTM network enables a high degree of generalization while this study specifically targets heart disease detection.

**Applicability Across Datasets:** The network has shown potential to be adaptable across varied ECG datasets, irrespective of their sources, making it a universally relevant model.

**Consideration of Varied Heart Conditions:** Though direct comparisons with other studies are intricate due to different conditions and test sets being considered, the general trend indicates a favorable skew towards our method when adjusted for these variances.

**Inclusion of Rare and Complex Conditions:** The network's depth and complexity allow it to detect even the rarer heart conditions, which often escape more rudimentary analytical tools.

#### C. Advantages of the Proposed Network

The 3D deep Conv-LSTM network brings a myriad of benefits to the table:

**Depth and Precision:** Leveraging the depth of convolutional neural networks (CNNs) and the sequential data handling ability of Long Short Term Memory (LSTM) networks, the model achieves an intricate blend of feature extraction and sequential data analysis [44]. This leads to nuanced detections that might be overlooked by shallow models.

**Reduced Overfitting:** The combination of CNN and LSTM, when architected correctly, curtails the typical problem of overfitting seen in deep networks [45]. This ensures that the model remains robust and versatile across varied datasets.

**Efficient Handling of Time-Series Data:** ECG data is inherently sequential, and the LSTM component of the network is adept at managing such time-series data, ensuring that temporal patterns, critical for heart disease detection, aren't missed [46].

**Scalability:** Given the rising volumes of ECG data, scalability is paramount. The proposed network, due to its architecture, is scalable both in terms of data volume and computational complexity.

#### D. Comparison with Previous Research

While earlier research primarily revolved around feature-based machine learning or shallow neural networks, the introduction of the 3D deep Conv-LSTM network marks a shift towards more intricate, end-to-end learning models [47]. It amalgamates the spatial feature learning capabilities of CNNs with the temporal sequencing prowess of LSTMs, making it a comprehensive solution.

#### E. Future Implications

As healthcare increasingly embraces technology, the proposed network offers a promising pathway for:

**Integrated Healthcare Systems:** ECG monitoring can be embedded into broader health monitoring systems, allowing for holistic health evaluations.

**Personalized Patient Care:** With a high degree of accuracy and early detection capabilities, treatments can be tailored based on the individual nuances detected by the network [48].

**Telemedicine and Remote Monitoring:** The network can be deployed in remote patient monitoring systems, democratizing access to quality cardiac care and reducing the need for frequent hospital visits [49].

In conclusion, the 3D deep Conv-LSTM network, as proposed, encapsulates the advancements and potentialities in the field of ECG-based heart disease detection. Its depth, versatility, and high accuracy make it a formidable tool in the evolving landscape of cardiac care. As data volumes grow and healthcare needs become more intricate, such networks will play a pivotal role in shaping the future of cardiac diagnostics and treatments.

## VI. CONCLUSION

In the ever-evolving landscape of cardiological research and diagnostics, the incorporation of advanced computational methodologies stands out as a quintessential advancement. This study delved into the efficacy of the 3D deep Conv-LSTM network, shedding light on its potential as a formidable tool for ECG-based heart disease detection. As the results elucidate, this proposed model not only embodies the intricate blend of spatial and temporal data handling but also surpasses the conventional methods in terms of precision and versatility.

The nexus between cardiology and computational modeling, especially as epitomized by the deep Conv-LSTM network, emphasizes the paradigm shift from rudimentary detection techniques to sophisticated, data-driven approaches. Our findings accentuate the network's capability to provide nuanced insights, thereby facilitating the early detection of

cardiac anomalies, including those that are often elusive in traditional assessments. Such advancements, as this research suggests, are imperative in the face of growing cardiac health challenges and the increasing need for preventive healthcare strategies.

Moreover, the broader implications of this research transcend its immediate findings. The proposed network's scalability and adaptability indicate its potential for integration into holistic healthcare systems, potentially revolutionizing patient care by offering tailored treatments and reducing the necessity for invasive procedures. Furthermore, as the realms of telemedicine and remote patient monitoring burgeon, models like the 3D deep Conv-LSTM can be pivotal in democratizing quality cardiac care, irrespective of geographical and infrastructural constraints.

In summation, this exploration into the 3D deep Conv-LSTM network underscores the confluence of cardiology and advanced computational methods. The ensuing synergies not only promise enhanced diagnostic capabilities but also chart the course for future research, emphasizing the inexorable march of technology in augmenting healthcare outcomes. The journey from ECG data acquisition to actionable cardiac insights, as portrayed by this study, is both a testament to current scientific progress and a beacon for future endeavors.

#### ACKNOWLEDGMENT

This work was supported by the research project "Application of machine learning methods for early diagnosis of Pathologies of the cardiovascular system" funded by the Ministry of Science and Higher Education of the Republic of Kazakhstan. Grant No. IRN AP13068289.

#### REFERENCES

- [1] Midani, W., Ouarda, W., & Ayed, M. B. (2023). DeepArr: An investigative tool for arrhythmia detection using a contextual deep neural network from electrocardiograms (ECG) signals. *Biomedical Signal Processing and Control*, 85, 104954.
- [2] Kumar, A. K., Ritam, M., Han, L., Guo, S., & Chandra, R. (2022). Deep learning for predicting respiratory rate from biosignals. *Computers in biology and medicine*, 144, 105338.
- [3] Malakouti, S. M. (2023). Heart disease classification based on ECG using machine learning models. *Biomedical Signal Processing and Control*, 84, 104796.
- [4] Tursynova, A., & Omarov, B. (2021, November). 3D U-Net for brain stroke lesion segmentation on ISLES 2018 dataset. In 2021 16th International Conference on Electronics Computer and Computation (ICECCO) (pp. 1-4). IEEE.
- [5] Singh, P., Sharma, A., & Maiya, S. (2023). Automated atrial fibrillation classification based on denoising stacked autoencoder and optimized deep network. *Expert Systems with Applications*, 233, 120975.
- [6] Dissanayake, T., Fernando, T., Denman, S., Sridharan, S., & Fookes, C. (2023). DConv-LSTM-Net: A Novel Architecture for Single and 12-Lead ECG Anomaly Detection. *IEEE Sensors Journal*.
- [7] Zhao, X., Yan, H., Hu, Z., & Du, D. (2022). Deep spatio-temporal sparse decomposition for trend prediction and anomaly detection in cardiac electrical conduction. *IIEE Transactions on Healthcare Systems Engineering*, 12(2), 150-164.
- [8] Shaqiri, E., Gusev, M., Puposka, L., Vavlukis, M., & Ahmeti, I. (2021, September). Comparing Time and Frequency Domain Heart Rate Variability for Deep Learning-Based Glucose Detection. In *International Conference on ICT Innovations* (pp. 188-197). Cham: Springer International Publishing.
- [9] Saripuddin, M., Suliman, A., Syarmila Sameon, S., & Jorgensen, B. N. (2021, September). Random undersampling on imbalance time series data for anomaly detection. In *Proceedings of the 2021 4th International Conference on Machine Learning and Machine Intelligence* (pp. 151-156).
- [10] Pan, J., & Tompkins, W. J. (1985). A real-time QRS detection algorithm. *IEEE transactions on biomedical engineering*, (3), 230-236.
- [11] Acharya, U. R., Oh, S. L., Hagiwara, Y., Tan, J. H., & Adam, M. (2017). Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. *Computers in biology and medicine*, 100, 270-278.
- [12] Prakarsha, K. R., & Sharma, G. (2022). Time series signal forecasting using artificial neural networks: An application on ECG signal. *Biomedical Signal Processing and Control*, 76, 103705.
- [13] 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.
- [14] Witt, D. R., Kellogg, R. A., Snyder, M. P., & Dunn, J. (2019). Windows into human health through wearables data analytics. *Current opinion in biomedical engineering*, 9, 28-46.
- [15] Witt, D. R., Kellogg, R. A., Snyder, M. P., & Dunn, J. (2019). Windows into human health through wearables data analytics. *Current opinion in biomedical engineering*, 9, 28-46.
- [16] Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1492-1500).
- [17] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory [J]. *Neural computation*, 9(8), 1735-1780.
- [18] Kiranyaz, S., Ince, T., & Gabbouj, M. (2016). Real-time patient-specific ECG classification by 1-D convolutional neural networks. *IEEE Transactions on Biomedical Engineering*, 63(3), 664-675. ↵
- [19] Rajpurkar, P., Hannun, A. Y., Haghpanahi, M., Bourn, C., & Ng, A. Y. (2017). Cardiologist-level arrhythmia detection with convolutional neural networks. *arXiv preprint arXiv:1707.01836*. ↵
- [20] Shi, X., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., & Woo, W. C. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In *Advances in neural information processing systems* (pp. 802-810).
- [21] Lu, Y., Wang, H., Zhou, B., Wei, C., & Xu, S. (2022). Continuous and simultaneous estimation of lower limb multi-joint angles from sEMG signals based on stacked convolutional and LSTM models. *Expert Systems with Applications*, 203, 117340.
- [22] Jafari, M., Shoeibi, A., Khodatars, M., Ghassemi, N., Moridian, P., Alizadehsani, R., ... & Acharya, U. R. (2023). Automated diagnosis of cardiovascular diseases from cardiac magnetic resonance imaging using deep learning models: A review. *Computers in Biology and Medicine*, 106998.
- [23] Christabel, G. J., & Subhajini, A. C. (2023). KPCA-WRF-prediction of heart rate using deep feature fusion and machine learning classification with tuned weighted hyper-parameter. *Network: Computation in Neural Systems*, 1-32.
- [24] Christabel, G. J., & Subhajini, A. C. (2023). KPCA-WRF-prediction of heart rate using deep feature fusion and machine learning classification with tuned weighted hyper-parameter. *Network: Computation in Neural Systems*, 1-32.
- [25] Liu, X., Shen, Y., Zhang, S., & Zhao, X. (2018). Segmentation of left atrium through combination of deep convolutional and recurrent neural networks. *Journal of Medical Imaging and Health Informatics*, 8(8), 1578-1584.
- [26] Liu, Z., Alavi, A., Li, M., & Zhang, X. (2023). Self-Supervised Contrastive Learning for Medical Time Series: A Systematic Review. *Sensors*, 23(9), 4221.
- [27] Yang, T., Li, G., Yuan, S., Qi, Y., Yu, X., & Han, Q. (2023). The LST-SATM-net: A new deep feature learning framework for aero-engine hydraulic pipeline systems intelligent faults diagnosis. *Applied Acoustics*, 210, 109436.



- [28] Banerjee, S., & Singh, G. K. (2023). A new real-time lossless data compression algorithm for ECG and PPG signals. *Biomedical Signal Processing and Control*, 79, 104127.
- [29] Wang, X., Zhang, S., Song, J., Liu, Y., & Lu, S. (2023). Magnetic signal denoising based on auxiliary sensor array and deep noise reconstruction. *Engineering Applications of Artificial Intelligence*, 125, 106713.
- [30] Li, D., Liu, J., & Zhao, Y. (2022). Prediction of Multi-Site PM2.5 Concentrations in Beijing Using CNN-Bi LSTM with CBAM. *Atmosphere*, 13(10), 1719.
- [31] Kareem, K. Y., Seong, Y., Bastola, S., & Jung, Y. (2022). Current State of Deep Learning Application to Water-related Disaster Management in Developing Countries. *Natural Hazards and Earth System Sciences Discussions*, 1-33.
- [32] Pieszko, K., Shanbhag, A., Killekar, A., Miller, R. J., Lemley, M., Otaki, Y., ... & Slomka, P. J. (2023). Deep learning of coronary calcium scores from PET/CT attenuation maps accurately predicts adverse cardiovascular events. *Cardiovascular Imaging*, 16(5), 675-687.
- [33] Zeng, Q., Liang, Y., Chen, G., Duan, H., & Wu, Q. (2023). A Novel Long-Term Noise Prediction System Based on  $\alpha$  DTW-DCRNN Using Periodically Unaligned Spatiotemporal Distribution Sequences. *IEEE Systems Journal*.
- [34] Martín-Escudero, P., Cabanas, A. M., Dotor-Castilla, M. L., Galindo-Canales, M., Miguel-Tobal, F., Fernández-Pérez, C., ... & Giannetti, R. (2023). Are Activity Wrist-Worn Devices Accurate for Determining Heart Rate during Intense Exercise?. *Bioengineering*, 10(2), 254.
- [35] Zucker, E. J. (2022). Compact pediatric cardiac magnetic resonance imaging protocols. *Pediatric Radiology*, 1-16.
- [36] 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.
- [37] Gonsalves, A. H., Thabtah, F., Mohammad, R. M. A., & Singh, G. (2019, July). Prediction of coronary heart disease using machine learning: an experimental analysis. In *Proceedings of the 2019 3rd International Conference on Deep Learning Technologies* (pp. 51-56).
- [38] Sharma, S., & Parmar, M. (2020). Heart diseases prediction using deep learning neural network model. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 9(3), 2244-2248.
- [39] Miao, K. H., & Miao, J. H. (2018). Coronary heart disease diagnosis using deep neural networks. *international journal of advanced computer science and applications*, 9(10).
- [40] Kumar, M. N., Koushik, K. V. S., & Deepak, K. (2018). Prediction of heart diseases using data mining and machine learning algorithms and tools. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 3(3), 887-898.
- [41] Bharti, R., Khamparia, A., Shabaz, M., Dhiman, G., Pande, S., & Singh, P. (2021). Prediction of heart disease using a combination of machine learning and deep learning. *Computational intelligence and neuroscience*, 2021.
- [42] 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.
- [43] Omarov, B., Tursynova, A., Postolache, O., Gamry, K., Batyrbekov, A., Aldeshov, S., ... & Shiyapov, K. (2022). Modified UNet Model for Brain Stroke Lesion Segmentation on Computed Tomography Images. *Computers, Materials & Continua*, 71(3).
- [44] Rath, A., Mishra, D., Panda, G., & Satapathy, S. C. (2021). Heart disease detection using deep learning methods from imbalanced ECG samples. *Biomedical Signal Processing and Control*, 68, 102820.
- [45] Omarov, B., Narynov, S., Zhumanov, Z., Gumar, A., & Khassanova, M. (2022). State-of-the-art violence detection techniques in video surveillance security systems: a systematic review. *PeerJ Computer Science*, 8, e920.
- [46] Yadav, S. S., Jadhav, S. M., Nagrale, S., & Patil, N. (2020, March). Application of machine learning for the detection of heart disease. In *2020 2nd international conference on innovative mechanisms for industry applications (ICIMIA)* (pp. 165-172). IEEE.
- [47] Ramesh, T. R., Lilhore, U. K., Poongodi, M., Simaiya, S., Kaur, A., & Hamdi, M. (2022). Predictive analysis of heart diseases with machine learning approaches. *Malaysian Journal of Computer Science*, 132-148.
- [48] Obasi, T., & Shafiq, M. O. (2019, December). Towards comparing and using Machine Learning techniques for detecting and predicting Heart Attack and Diseases. In *2019 IEEE international conference on big data (big data)* (pp. 2393-2402). IEEE.
- [49] Ghumbre, S. U., & Ghatol, A. A. (2012). Heart disease diagnosis using machine learning algorithm. In *Proceedings of the International Conference on Information Systems Design and Intelligent Applications 2012 (INDIA 2012)* held in Visakhapatnam, India, January 2012 (pp. 217-225). Springer Berlin Heidelberg.

# DetBERT: Enhancing Detection of Policy Violations for Voice Assistant Applications using BERT

Rawan Baalous, Joud Alzahrani, Mariam Ali, Rana Asiri, Eman Nooli  
Cybersecurity Department, University of Jeddah, Jeddah, Saudi Arabia

**Abstract**—Voice Assistants, also known as VAs, have gained popularity in the last few years. They make our daily tasks easier via simple voice instructions. VAs platforms allow third-party developers to develop voice applications and publish them on the VAs platforms. However, VAs applications may collect users' personal information for different purposes. To maintain the security and privacy of users, VAs platforms have specified a set of policies that must be adhered to by VAs applications' developers. This paper aims to automatically detect voice apps that do not comply with the VA's platforms policies. To this end, DetBERT, a comprehensive testing tool, was built. DetBERT evaluates voice apps' compliance with the policies using BERT model by analyzing the apps' behaviors and detecting violations. With DetBERT, a total of 50,000 voice assistant apps from Amazon Alexa and Google Assistant platforms were tested. The paper demonstrates that DetBERT can accurately identify whether a voice assistant application has violated the platform's policy or not.

**Keywords**—Alexa; Google assistant; BERT; policy violation detector; voice assistant; user privacy; security

## I. INTRODUCTION

Voice assistants (VAs) have become widespread and integrated into billions of people's daily lives due to their ease of everyday tasks and the comfortable services they provide. Amazon Alexa and Google Assistant are ones of the most popular VAs platforms, which allow third-party developers to publish their voice applications<sup>1</sup> in the stores [1]. Many users' activities can be accomplished through these skills, including placing orders, obtaining information like weather and news, and making phone calls. This attracts tens of millions of users and, in turn, more developers.

As skills grow rapidly, dangerous skills also appear. Since third-party developers can share their skills, the privacy and security concerns of VAs users arise regarding the skills developers' intents [2]. Recent studies have revealed that developers are capable of redirecting users' requests to malicious skills without their knowledge. This can be achieved by naming their skills similarly to legitimate ones [3][4]. In fact, malicious skills could eavesdrop on users' conversations or even monitor them, which affect users' privacy [5]. In order to maintain the security and privacy of users, VAs platforms have defined a set of policy requirements and enforced them to be adhered to by third-party developers. Nonetheless, some

VAs platforms use a weak vetting system [1], which allows several skills that violate policies to bypass the VAs platform's verification process and get certified.

The main challenge obstructing authoritative skill certification is the VAs platforms' distributed architecture. Using static code analysis to investigate a skill's behavior is not an option for current VAs systems. This is because the skill's code is hosted externally in the developer's servers, making it not accessible. As a result, the only way to comprehend a skill's actual behavior is through dynamic analysis (by interacting with a skill) [6]. This motivated us to explore VA's skills that violate policy requirements by behaving suspiciously, such as asking for personal information when it is not supposed to request such details.

In this work, we aim to identify stealthy policy violations conducted by third-party developers in Amazon Alexa and Google Assistant by enhancing the robustness and accuracy of the policy-violation detection process. Prior works showed many limitations in the approaches used by the pre-crafted policy-violation detection tools [6][7]. The drawback of these approaches lies in the inability to handle a skill's textual speech in a contextual meaning for the entire sentence. This results in decreasing the accuracy of detecting violations among variant policies type. To this end, we created VA's policy-violation detection tool using Bidirectional Encoder Representations Transformers (BERT) model. BERT [8] works in a bidirectional way to figure out the ambiguous language in the text, hence increasing the accuracy of analyzing the speech context.

In summary, this paper contributes to the field of privacy compliance checking by enhancing VA's policy violation detection. We developed a dynamic policy violation detection tool, called DetBERT. Our tool utilizes BERT model to improve the process of detecting skills that violate the VA's platforms policies. We developed two BERT models: User Privacy BERT and Content Safety BERT. The accuracy of both models in identifying violations is 0.98 % and 0.93 respectively.

The reminder of this paper is organized as follows: The background survey is detailed in Section II. Section III summarizes the recent literature on policy violation detection of VAs as well as several attacks on them. The process of detecting skills that violated the platforms policies is discussed in Section IV. Sections V and VI present the results of violations detected and compare the results with previous works. Finally, Section VII presents the conclusion, limitations and future work.

<sup>1</sup>Voice-apps are known as skills on Amazon and actions on Google. In this paper, we refer to voice applications as skills unless there is a need to clearly distinguish between the two platforms.

## II. BACKGROUND

### A. VAs Platforms and Skills Interaction

Fig. 1 presents an overview of the VA platform and skill interaction flow. A skill comprises a front-end interaction model and a back-end cloud service (skill code). The back-end is responsible for handling requests that come from users and directing a VA device's response. Similar to smartphone applications, most skills are created by third-party developers and made accessible through a skill store website. Skill's front-end and back-end are hosted separately due to the distributed architecture of VAs platforms. The back-end code is typically hosted on the developer's server (e.g., hosted by Amazon Web Services AWS Lambda under the developer's account or other third-party servers). On the other hand, the VAs platforms host the front-end interface [1].

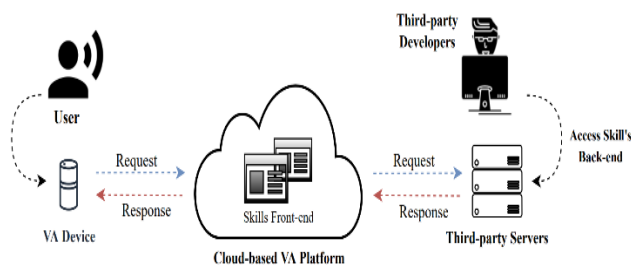


Fig. 1. Cloud-based alexa platform.

Amazon and Google provide an online repository of skills through their skills stores. Each skill is an individual product with a unique web page in the store. Skill's webpage includes the developer's information, skill description, skill identifier, privacy statement, users ratings, and users reviews [9]. In addition, it includes a sample of utterances (i.e., skill invocations) that enable the user to interact with the skill. A skill's privacy policy on the skill store should describe whatever data the skill may collect and how it will use that data in the future. Some skills may ask users to grant access to personal information to receive customized services. In this case, the user must provide permission through the VA companion app (Android/iOS) for the skill to get the required personal data [6]. VAs platforms provide a skill simulator for the testing needs of developers' skills [10]. The skill simulators provide a text-based interface that will receive text input, produce text output, and deliver external content to make testing more manageable. VA's simulators consist of a virtual VA device that can communicate with other skills available in the skill store [6].

### B. VAs Platforms Policies

VAs platforms provide a set of policy and security requirements that skills must adhere to. Before publishing a skill to the store, the platforms conduct a test to verify if the skill complies with these requirements. This process is designed to restrict the amount of potentially exploitable content that may be found on the skill store. If the VAs platform determines the existence of a violation, it has the right to take disciplinary action. Amazon Alexa specified 14 policy requirements and 7 privacy requirements [11][12]. Under each requirement, there are several statements that

describe illegal skill usage regarding the policy. On the other hand, Google Assistant has 10 main sections of policies [13]. Every section covers several policies related to it. To demonstrate, Google has a section called Content Restrictions that specifies allowed and disallowed content. This section involves 15 policies related to content. Another section is Privacy and Security which establishes requirements of what data is allowed to be collected, how skills must handle users' data, and maintain security among skills. There are common policies between Amazon and Google regarding specific skills categories (e.g., Health and Kids).

### C. BERT

In late 2018, researchers at Google AI developed a state-of-art model that has been an inflection point for Natural Language Processing (NLP). **B**idirectional **E**ncoder **R**epresentations from **T**ransformers is a machine learning pre-trained model. BERT has been pre-trained on a large dataset of books (800 million words) and texts from English Wikipedia pages (2,500 million words). It combines the right-left and left-right contexts to create a complete picture of the text. As a result, it helps machines to understand the contextual relation between words in the sentence. BERT uses transfer learning which improves the fine-tuning-based approaches. This means training a model on a general task (pre-training), then taking advantage of the knowledge that the model gained to solve related tasks (fine-tuning). With just one output layer, the BERT model can be fine-tuned to produce state-of-the-art text representations for various tasks, such as question answering, semantic analysis, and text prediction [8].

There are two main versions of BERT model: BERT base and BERT large. The main difference between the two versions is the number of used encoders. The first version: BERT base consists of 12 encoders, whereas the second version: BERT large consists of 24 encoders that originate from the transformer model. The large version of BERT represents more robust than the BERT base version, therefore it requires more powerful resources [8].

BERT uses special tokens as part of its input representation. These special tokens serve specific functions in the BERT input format and help the model process the input text correctly. Positions and meanings of the special tokens are taken into consideration by the model when generating representations for the input tokens. The most commonly used special tokens in BERT are [CLS], and [SEP]. The [CLS] or Classification token is added at the beginning of the input sequence and is used to represent the entire input sequence for classification tasks. In other words, the output of the BERT model for the [CLS] token is used as a representation of the entire input text for classification tasks, such as sentiment analysis or named entity recognition. The second token is the [SEP] or (separation) token. It is added at the end of the first sentence and in between subsequent sentences in the input sequence. It is used to separate the different sentences in the input text. This allows BERT to treat each sentence independently and capture the relationship between the sentences, which is important for tasks like question answering [8].

### III. RELATED WORKS

In this section, we summarize recent literature on policy violation detection of VAs as well as several attacks on them.

With regard to policy violation detection, Cheng et al. [1] conducted a study about the trustworthiness of skill certification in Amazon Alexa and Google Assistant platforms in terms of catching any policy violations in the third-party skills and whether policy-violating skills are published in the stores. The authors were interested in evaluating the level of difficulty in publishing skills that violate the policies in the stores. Intentionally, they submitted 234 Alexa skills and 381 Google Assistant actions that violated privacy and content policies. Surprisingly, all the violated skills got certified. At the end, the authors provided strategies in order to enhance the skills certification process. Similarly, Lentzsch et al. [14] identified flaws in the vetting process conducted by Amazon and tested only the skills that request permissions for data collection.

Dynamic testing tools [6][7][15][16] have been developed to enable automated skills analysis on a broad scale. SkillDetective [6] is a testing tool that explores voice apps' behaviors and identifies possible policy violations through live interaction with skills. The authors tested 54,055 Amazon Alexa skills and 5,583 Google Assistant actions and identified 6,079 skills and 175 actions violating at least one policy requirement. They utilized data-driven methodology to identify question types and used the Feedforward Neural Networks (FNN) and a bag-of-words (BoW) approach for answer prediction. However, implementing FNN and a BoW approach in the SkillDetective Chatbot slowed the policy-violation detection process. As a result, SkillDetective could not test every skill at full capacity. On the other hand, Guo et al. [7] developed SkillExplorer, an automated testing tool used to examine a skill's behavior through a grammar-rule based technique. The tool mainly focused on skills that collect private information from users. Authors found that 1,141 skills and 1,897 actions request personal information from users without specifying that in their privacy policies. As a consequence of using grammar-rule based technique, SkillExplorer was not able to properly answer some questions during the interaction with skills. This affected the accuracy of the violation detection results negatively.

Focusing on health VAs, Shezan et al. [15] proposed a static and dynamic machine learning-based solution. The dynamic part triggered and detected the violation through deep interaction with 813 health-related skills in Alexa. At the same time, the static part analyzed the web page of these skills. The study aimed to detect skills that provide life-saving assistance and lack disclaimers, which is prohibited by Amazon. They consulted medical school students regarding the correctness of the potential violation. In the end, VerHealth detected 244 out of 813 skills violating Amazon's health-related policies. In terms of kids related apps, authors in SkillBot [16] developed an automatic tool using natural language processing that interacts only with kids-related skills. They aimed to find out the risky skills that may ask for kids' personal information or contain inappropriate content. The tool analyzed 3,434 Alexa kids skills. Results showed that there are 28 risky child-

directed skills. In addition, a user-study has been conducted with parents. The authors also identified a novel risk in VAs called, confounding utterance. They defined it as: voice commands that invoke a non-child skill over a child-directed skill. In the end, they found 4,487 confounding utterances which indicate the high risk surrounding the children of invoking a non-child skill by accident.

Considering attacks on VAs, Cheng et al. [17] proved that skills are features that provide an entry point for attackers. They analyzed multiple attacks on VAs. Also, there are many researches on hidden voice attacks [18]–[20] and their corresponding defenses [21][22]. Kumar et al. [23] validated skill squatting attacks. In this attack, the attacker gets the advantage of sentence ambiguities and similar pronunciation to redirect the VAs users into a malicious skill. There are many types of masquerading attack, including voice squatting, in which the attacker exploits how the user call the skill and alter this call either by imitating the skill call or reordering the sentence words. In voice masquerading, the malicious skill impersonates the VA service to gather users' personal information or eavesdrops on their private conversations. Richard et al. [24] showed man-in-the-middle attacks against benign skills. The attack utilizes a weakness presented in a skill interface to redirect a victim's voice when invoking the skill to a malicious skill, then hijack the conversation between Alexa and the victim.

### IV. METHODOLOGY

This section provides an overview of the whole process of detecting skills that violated the platforms' policies (Section A). After that, the key modules of the process are discussed in detail. The data collection procedure is firstly presented (Section B). Then, the interaction with the skills procedure (Section C). Lastly, the violation detection procedure using the BERT model (Section D).

#### A. Overview

As illustrated in Fig. 2, the first step in the policy violation detection process is the interaction with a skill. To analyze a skill's potentiality of violating the policy, the skill's outputs to users must be collected. To this end, a chatbot that communicates with the VA's device simulator has been used. The chatbot automatically interacts with the targeted skill to collect its outputs, making the process faster and easier than manual interaction. The interaction starts when the first utterance (e.g., "Alexa, Open My Nutrition") is fed to the skill, resulting in activating it. When the skill receives an invocation word, it will pass back an output. During the communication, all skill's outputs will be stored to be examined later for policy violation. When the communication with the simulator ends, the violation detection process will be started in offline mode. The collected outputs are passed to the policy violation detector tool which analyzes the gathered outputs to identify any violation. To accomplish this process, the tool utilizes the BERT model that is trained on analyzing and classifying the outputs, searching for violation indications. Using BERT in the detector tool helps to understand the ambiguous violation in the sentences and phrases. As a result, it improves the accuracy of policy violation detection in the skill's outputs.

### B. Data Collection

1) *Skills sample*: As a primary dataset, we used SkillDetective's dataset [6]. Each record in the dataset consists of the skill's data divided into six features. Table I describes each feature in the dataset. However, as the store releases new skills continuously, and to ensure violation detection of recently published skills, we have collected more skills using the Octuparse extraction tool [25]. Using this tool, we can automatically access web pages and extracts data from them. Once we have finished the extraction process, we combined our skills dataset and SkillDetective's dataset, with paying attention to remove duplicated skills. In total, the final skills sample is 69,843 skills and 16,003 actions.

2) *Violations dataset*: We had many challenges while looking for suitable datasets to fine-tune BERT Model on violation detection based on Amazon Alexa and Google Assistant policies. The reason is that BERT receives datasets in form of sentences and paragraphs (i.e., Tweets). The lack of User Privacy Violations Datasets led us to craft one from scratch. For other types of violations, we used publicly published datasets.

TABLE I. FEATURES DESCRIPTION IN SKILLDETECTIVE'S DATASETS

Feature	Description
Skill ID	Unique identifier of the skill, consists of 10 digits.
Name	Specifies skill name in the store.
Category	Specifies which categories the skill belongs to.
Invocation	Keywords used to start the interaction with skill.
Description	Specifies the skill functionalities.
Privacy Policy Link	Specifies the privacy policy that the skill adheres to.

The User Privacy Violation Dataset consists of sentences and questions mentioned during the conversation, with the

intent of collecting information about users. For example, Are you alone at home? or Provide me with your graduation date. For help in creating this dataset, we used ChatGPT [25], which is an AI-powered language model that generates human-like responses to various questions and prompts. Based on the VAs platforms policies, we prompted ChatGPT to generate questions and sentences that violate user privacy. We specifically asked for queries related to collecting personal, sensitive, health, and kids' information, since the privacy policies of the VAs state different permissions regarding collecting user information based on skill's category. In that manner, our dataset contains four classes: 1) Personal Data Collection, 2) Sensitive Data Collection, 3) Health Data Collection, and 4) Non-violation. Table II presents some examples of information involved in each class. Many of the generated questions were duplicated, so we had to remove duplication using python scripts. The total number of questions we got was 3745 unique questions. After collection, the second step was the annotation. We chose a human annotation approach because it is often more accurate compared to automated approaches. We manually checked for the mismatch between data contents and their corresponding labels. The process was iterative across the authors.

The Content Safety Violation Dataset consists of content identified as harmful and inappropriate in policies (e.g., sexual content). Such content is prohibited from occurring or being mentioned by the skill. We used two published datasets from Kaggle, Toxic Content [26] and Cyberbullying datasets [27]. They consist of two columns: Toxicity (toxic and non-toxic), and Content. We merged both datasets for fine-tuning, resulting in 207,266 records.

### C. Interaction with the Skills

Once the interaction with the skill starts, the goal is to maintain the conversation continued between the chatbot and the skill as long as possible. By keeping the conversation ongoing, we can gather more skills' outputs to be analyzed. As a result, more skills' behaviors will be identified. To this end, we used the SkillDetective Chatbot [28], which was published for future research.

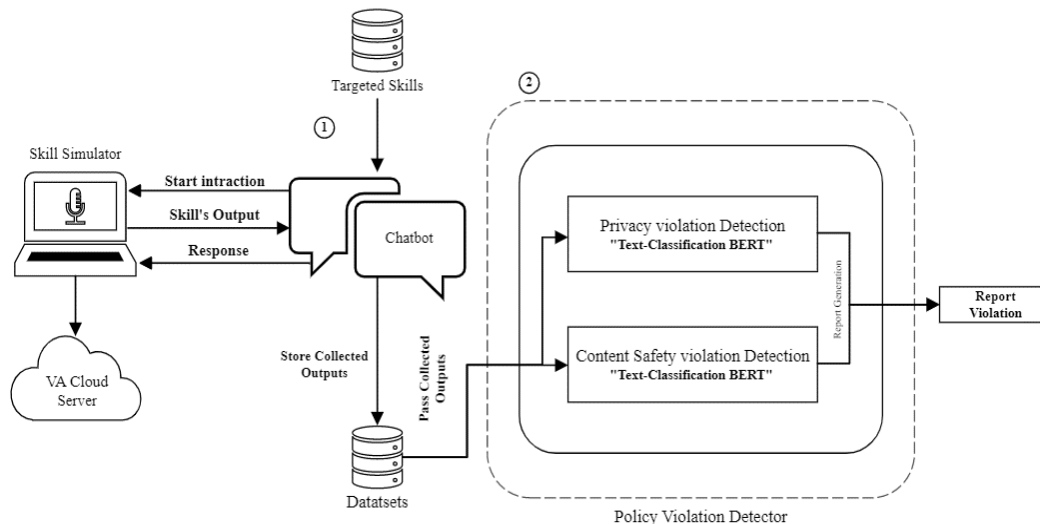


Fig. 2. DetBERT methodology.

Of all the published chatbots that previous works have used, we chose SkillDetective Chatbot for two reasons. First, the chatbot can handle five types of questions with high classification accuracy, as shown in Table III. The second reason is that it implements an approach called SkillTree. This approach is meant to build a dynamically growing tree to keep track of all branches (i.e., paths) that have been examined and those that have not, to ensure all the skill's possible behaviors have been explored. It is used in situations where questions with multiple answers exist. For example, the Yes/No questions have two answers, and the selection questions have two or more answers. Each answer is saved in the branch of the tree to be visited later. Using this approach helps to increase skill's coverage and reduces testing latency. One drawback of the chatbot, it utilizes FNN and BoW approaches, which make communication heavy and very slow. Besides, the continuous interruptions of the connection with the simulator which required human intervention. Lastly, by the end of every interaction, all outputs generated by the skill are recorded for later analysis.

D. Violation Detection

The policy violation detector tool focuses on detecting violations that happened during the interaction with skill. In this work, we mainly focus on two types of policies: 1) User Privacy Policies; 2) Content Safety Policies. These policies are described in details in the Appendix (Table VIII), which lists the policy statements as mentioned in the VAs platforms. To detect violations related to these policies, we used BERT base model to build our tool. We developed two different BERT models for the text classification task according to the different violations we looked for during the examination. The first model is a multi-classification model which was developed to check for User Privacy Violations. The second model is a binary-classification model trained on detecting Content Safety Violations. We mainly took pre-trained BERT models and then fine-tuned them on specific datasets to lower the cost of BERT training. Using BERT in violation detection allows the sentences (skills' outputs) to be processed in a contextualized meaning for the entire output sentences. Therefore, the accuracy of detecting violations increases.

Fig. 3 illustrates how the policy violation detector works. First, the skill's output is tokenized using the wordPiece tokenizer [29], which is the BERT tokenizer. The tokenizer breaks down the sentence into chunks of words. Depending on the vocabulary file utilized by the tokenizer, some words are split into a single word, and some are split into multiple words. After tokenization, BERT special tokens, [CLS] and [SEP], are added at the beginning and end of the sentence. Then, the tokens are padded with the "PAD" token to reach the maximum input size for BERT. After padding the tokens, they are converted to token IDs which are fed to the BERT model. In the following step, the input tokens are transformed into integer IDs based on the tokenizer vocabulary file. These integer IDs are input into the BERT model along with a matrix of ones and zeros called an Attention Mask. The matrix represents whether the token input is genuine or padded input. The attention mask elements corresponding to genuine inputs are set to 1, while those corresponding to padded inputs are set to 0. The BERT model converts each genuine input into a

vector of a specific size known as the BERT hidden size. This vector is created by an encoder with an attention layer, which allows it to better understand the token's context. After all the vectors have been made, they are then used as input to a classification layer, which determines the class that the input belongs to.

TABLE II. EXAMPLES OF INFORMATION INVOLVED IN DATASET CLASSES

Class	Involved Information
Personal information	Name, Age, Birthday, Gender, Location
Sensitive data	Social Security Number, Visa Code, Passwords, Bank Account Numbers, Credit Card Numbers
Health data	Heart Rate, Mass Index, Blood Pressure, Blood Type
Non-violation	General Questions and Statements

TABLE III. SKILLDETECTIVE CHATBOT'S QUESTIONS CLASSIFICATION ACCURACY [6]

Question Type	Example	Identification Accuracy
Binary	Are you in the car?	100%
Selection	Say your name.	99.3%
Instruction	Do you want to eat, run, or watch TV?	99%
Open-Ended	What is your mother's name?	98%
Mixed	There are A, B, and C to choose from. Which one do you want?	98%

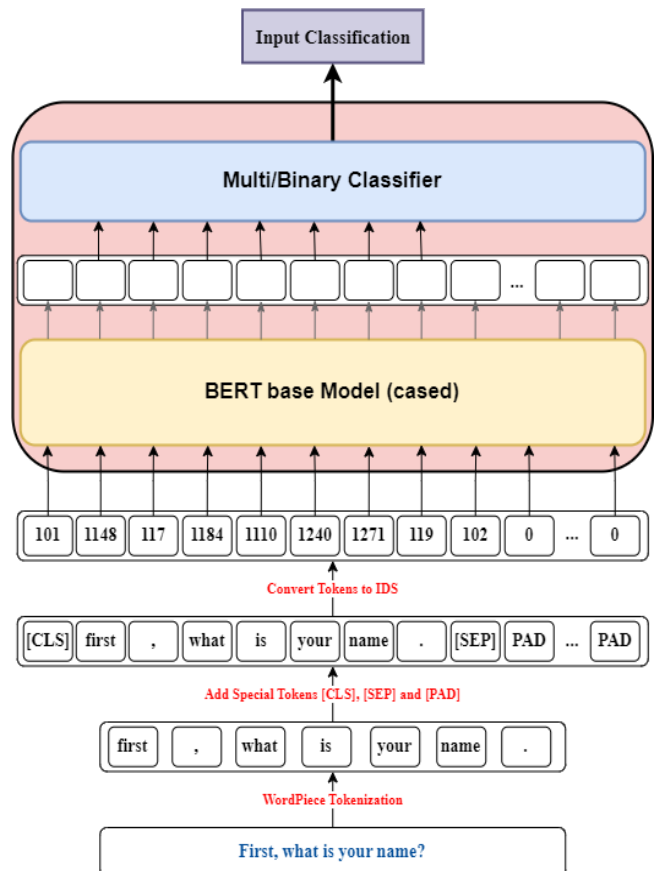


Fig. 3. How policy violation detector works.

The multi-classification BERT model which was developed to check for User Privacy Violations has been used to determine the violation type. It classifies each tweet (i.e. skill's output) entered the model into one of four classes, 1) Personal Data Collection; 2) Sensitive Data Collection; 3) Health Data Collection; 4) Non-violation. Note that skills are allowed to ask for personal data in order to perform some of their required tasks, with the condition of providing a privacy policy link outlining legal data usage. It is also important to highlight that skills related to health deal with sensitive data about users' medical conditions. These kinds of data cannot be collected or disclosed without user's permission. In a like manner, kids' skills are designed for kids who are targeted by potential threats more than adults. The tool detects any potential unauthorized data collection of users' personal or sensitive data. For personal data collection, the skill must attach a privacy notice (i.e., privacy policy link) to its page. The tool uses the User Privacy BERT model to determine whether the skill gathered user data during the conversation. If it does, the existence of privacy notices on the skill's page will be checked. If there is no link provided, the skill is considered violated.

On the other hand, Content Violations are hard to predict. In fact, the skill's behavior may differ based on the conversation user had with the skill or because of the skill's update. Both VAs platforms stated many policies related to content. Based on these policies, the policy violation detector used Content Safety BERT model to detect skills that violate content restrictions. All skills are not allowed to use inappropriate and harmful content like profanity or hate language and promoting or sale of illegal materials like drugs. For kids' skills, content must be appropriate for all ages. To this end, the binary BERT model is deployed to recognize and differentiate between harmful and legal content. It classifies each tweet entered in the model into one of two classes: 1) Toxic; 2) Non-Toxic.

To generate the final report, we summarized all the violations results after the detection process ends. The violations detected were recorded and saved into four files. Files were divided based on four categories of violations. The categories are as follows: 1) Kids Policy Violations; 2) Health Policy Violations; 3) Toxic Policy Violations; 4) General Policy Violations. We have divided the results into these files as they are the main types of violations we have focused on. Each file contains six primary columns: Order, Category, Violated\_policy, Skill\_id, Skill\_name, and Skill\_output.

### V. RESULTS

The results of violations detected in terms of user privacy violations and content safety are summarized in Table IV.

Table V presents the results obtained by our models: User Privacy BERT and Content Safety BERT, in terms of precision, recall, F1-score, and accuracy. The results indicate that User Privacy BERT model is performing extremely well in detecting policy violations related to user privacy. With the achieved accuracy of %98, the model makes correct predictions about violations. Overall, the results shown by the User Privacy BERT model demonstrate its high performance. On the other hand, the results obtained by the Content Safety

BERT model shows that the model is performing well, but there is still room for improvement. In fact, the model was trained on seven epochs, which is better to be increased especially for huge datasets like the content safety violations dataset.

TABLE IV. SUMMARY OF VIOLATIONS DETECTED BY DETBERT

Detected Violations Type		# Of skills	# Of actions
Violation of User Privacy Policies	Collect health data	147	0
	Collect personal data	52	14
	Collect sensitive data	0	0
	Lack of privacy policy	172	6
Violation of Content Safety	Contain toxic content in general categories	154	0
	Contain toxic content in kids category	1	0

TABLE V. PERFORMANCE EVALUATION OF DETBERT MODELS

Model	Accuracy	Precision	Recall	F1
User Privacy BERT	0.98	0.98	0.98	0.98
Content Safety BERT	0.93	0.93	0.93	0.93

Fig. 4 displays the User Privacy BERT model confusion matrix. The TP and TN show a high score, which in turn indicates the good performance of the model. The matrix reveals that the model has identified 1106 violations correctly out of 1125 actual values. Additionally, it accurately categorized 356 non-violations out of 375 actual values. The results of the confusion matrix of the Content Safety BERT model are shown in Fig. 5. The high score for TP and TN indicates also the model's good performance. The matrix shows that the model correctly identified 27,519 Toxic content violations out of 30,000. Additionally, it accurately categorized 28,018 Non-toxic contents out of 30,000.

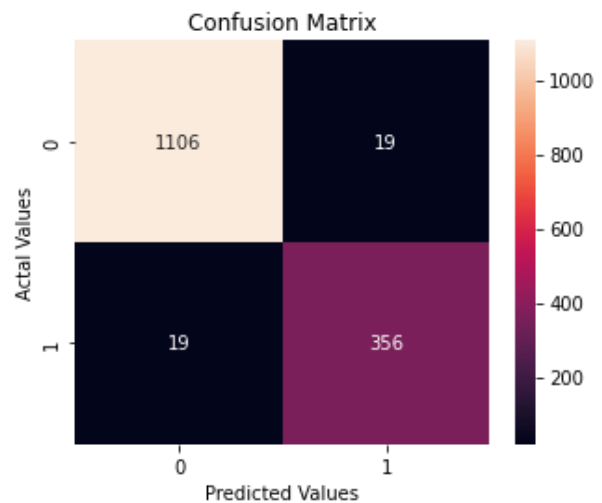


Fig. 4. Confusion matrix of user privacy violations dataset.

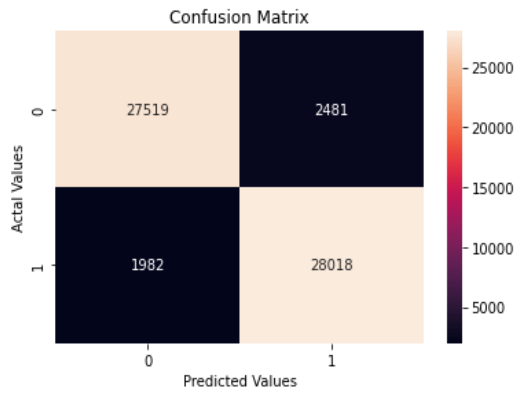


Fig. 5. Confusion matrix of content safety violations dataset.

In summary, we proved the high performance of our proposed DetBERT tool. This is done by utilizing confusion matrix analysis for BERT models. The results of both matrices showed successful TP and TN predictions by the models.

### VI. DISCUSSION

In order to compare our work with previous works, we conducted a performance comparison based on the results of the detected violations between DetBERT and SkillDetective[6], as shown in Table VI. The comparison was performed on the same sample for both models. The results of SkillDetective showed that 557 skills and 13 actions violated at least one policy. To ensure the correctness of the results, we conducted a manual revision of SkillDetective results. In fact, we have found some FP results in Content Safety. After excluding the FP, the final results obtained was 159 violations detected. On the other hand, The BERT-based approach of DetBERT has led to better performance in detecting policy violations as shown in the table. This comparison demonstrates the efficiency of DetBERT and its potential to provide valuable insights into policy compliance.

In terms of accuracy, Table VII shows the differences in accuracy between SkillDetective and DetBERT. Regarding user privacy policies, SkillDetective has developed two different methods to detect skills that collect user data. As summarized in Table VII, both methods achieved results less than what the User Privacy BERT model achieved. This comparison concludes that BERT-based model provided more accurate results in terms of detecting data collection policy violations. For content safety, SkillDetective did not reveal much details about accuracy results, and hence no comparison was provided.

TABLE VI. NUMBER OF VIOLATIONS DETECTED IN SKILLDETECTIVE AND DETBERT

Policy Violation Type	SkillDetective		DetBERT	
	# Of skills	# Of actions	# Of skills	#Of actions
User Privacy Policies	364	13	371	20
Content Safety Policies	159	0	171	0
Total Violated Skills	523	13	542	20

TABLE VII. ACCURACY COMPARISON BETWEEN SKILLDETECTIVE AND DETBERT

Tool	Model/Method	Accuracy
SkillDetective	Kids and Health Categories	92%
	Data Collection for general Categories	89%
DetBERT	User Privacy BERT	<b>98%</b>

### VII. CONCLUSION

To conclude, the detection of policies' violations is an ongoing challenging process and an open area for researchers in the NLP field. In this paper, we have presented an improvement in examining voice assistants apps' compliance with stores' policies using the BERT model. We designed and implemented a violation detection tool called DetBERT. The results provide valuable insights about policy violations and can be used in the future for more accurate policy compliance checking.

This research has some limitations. Due to a limitation in the server we used to run DetBERT, we could only test a total of 50,000 skills. In addition, there was no publicly available dataset related to policy violations for the voice assistants platforms. As a result, we ended up creating our dataset with manual annotation. The dataset consists of 3745 records, which was considered small for fine-tuning the model properly. We believe there are various ways to improve and extend this study in the future. Future works can consider using larger dataset for fine tuning BERT. In addition, creating a chatbot using AI such as ChatGPT that can engage in a conversation with the skills, may enhances the question-answering precision and provides insights into new and ongoing risky behaviors.

### REFERENCES

- [1] L. Cheng, C. Wilson, S. Liao, J. Young, D. Dong, and H. Hu, "Dangerous skills got certified: Measuring the trustworthiness of skill certification in voice personal assistant platforms," in Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security, 2020, pp. 1699–1716.
- [2] D. Su, J. Liu, S. Zhu, X. Wang, and W. Wang, " 'Are you home alone?' 'Yes' Disclosing Security and Privacy Vulnerabilities in Alexa Skills," arXiv Prepr. arXiv:2010.10788, 2020.
- [3] N. Zhang, X. Mi, X. Feng, X. Wang, Y. Tian, and F. Qian, "Dangerous skills: Understanding and mitigating security risks of voice-controlled third-party functions on virtual personal assistant systems," in 2019 IEEE Symposium on Security and Privacy (SP), 2019, pp. 1381–1396.
- [4] Y. Zhang, L. Xu, A. Mendoza, G. Yang, P. Chinpruthiwong, and G. Gu, "Life after speech recognition: Fuzzing semantic misinterpretation for voice assistant applications," in Proc. of the Network and Distributed System Security Symposium (NDSS'19), 2019.
- [5] D. Kumar et al., "Skill squatting attacks on Amazon Alexa," in 27th USENIX security symposium (USENIX Security 18), 2018, pp. 33–47.
- [6] J. Young, S. Liao, L. Cheng, H. Hu, and H. Deng, "SkillDetective: Automated Policy-Violation detection of voice assistant applications in the wild," in USENIX Security Symposium, 2022.
- [7] Z. Guo, Z. Lin, P. Li, and K. Chen, "{SkillExplorer}: Understanding the Behavior of Skills in Large Scale," in 29th USENIX Security Symposium (USENIX Security 20), 2020, pp. 2649–2666.
- [8] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," arXiv Prepr. arXiv:1810.04805, 2018.



[9] S. Liao, C. Wilson, L. Cheng, H. Hu, and H. Deng, "Measuring the effectiveness of privacy policies for voice assistant applications," in Annual Computer Security Applications Conference, 2020, pp. 856–869.

[10] "Test with the Alexa Simulator | Alexa Skills Kit." <https://developer.amazon.com/en-US/docs/alexa/devconsole/alexa-simulator.html> (accessed Oct. 01, 2022).

[11] "Policy Testing | Alexa Skills Kit." <https://developer.amazon.com/fr-FR/docs/alexa/custom-skills/policy-testing-for-an-alexa-skill.html> (accessed Sep. 20, 2022).

[12] "Security Testing for an Alexa Skill | Alexa Skills Kit." <https://developer.amazon.com/en-US/docs/alexa/custom-skills/security-testing-for-an-alexa-skill.html#25-privacy-requirements> (accessed Nov. 14, 2022).

[13] "Policies for Actions on Google | Actions console | Google Developers." <https://developers.google.com/assistant/console/policies/general-policies> (accessed Oct. 28, 2022).

[14] C. Lentzsch, S. J. Shah, B. Andow, M. Degeling, A. Das, and W. Enck, "Hey Alexa, is this skill safe?: Taking a closer look at the Alexa skill ecosystem," Netw. Distrib. Syst. Secur. Symp., 2021.

[15] F. H. Shezan, H. Hu, G. Wang, and Y. Tian, "Verhealth: Vetting medical voice applications through policy enforcement," Proc. ACM interactive, mobile, wearable ubiquitous Technol., vol. 4, no. 4, pp. 1–21, 2020.

[16] T. Le, D. Y. Huang, N. Apthorpe, and Y. Tian, "Skillbot: Identifying risky content for children in alexa skills," ACM Trans. Internet Technol., vol. 22, no. 3, pp. 1–31, 2022.

[17] P. Cheng and U. Roedig, "Personal Voice Assistant Security and Privacy—A Survey," Proc. IEEE, vol. 110, no. 4, pp. 476–507, 2022.

[18] N. Carlini et al., "Hidden voice commands," in 25th USENIX security symposium (USENIX security 16), 2016, pp. 513–530.

[19] Y. Wu et al., "HVAC: Evading Classifier-based Defenses in Hidden Voice Attacks," in Proceedings of the 2021 ACM Asia Conference on Computer and Communications Security, 2021, pp. 82–94.

[20] G. Zhang, C. Yan, X. Ji, T. Zhang, T. Zhang, and W. Xu, "Dolphinattack: Inaudible voice commands," in Proceedings of the 2017 ACM SIGSAC conference on computer and communications security, 2017, pp. 103–117.

[21] I.-Y. Kwak, J. H. Huh, S. T. Han, I. Kim, and J. Yoon, "Voice presentation attack detection through text-converted voice command analysis," in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 2019, pp. 1–12.

[22] C. Wang, S. A. Anand, J. Liu, P. Walker, Y. Chen, and N. Saxena, "Defeating hidden audio channel attacks on voice assistants via audio-induced surface vibrations," in Proceedings of the 35th Annual Computer Security Applications Conference, 2019, pp. 42–56.

[23] N. Zhang, X. Mi, X. Feng, X. Wang, Y. Tian, and F. Qian, "Understanding and mitigating the security risks of voice-controlled third-party skills on amazon alexa and google home," arXiv Prepr. arXiv1805.01525, 2018.

[24] R. Mitev, M. Miettinen, and A.-R. Sadeghi, "Alexa lied to me: Skill-based man-in-the-middle attacks on virtual assistants," in Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security, 2019, pp. 465–478.

[25] "ChatGPT: Optimizing Language Models for Dialogue." <https://openai.com/blog/chatgpt/> (accessed Feb. 13, 2023).

[26] "Toxic Comment Classification Challenge | Kaggle." <https://www.kaggle.com/competitions/jigsaw-toxic-comment-classification-challenge/data> (accessed Feb. 13, 2023).

[27] "Cyberbullying Classification | Kaggle." <https://www.kaggle.com/datasets/andrewmvd/cyberbullying-classification> (accessed Feb. 14, 2023).

[28] "skilldetective/ChatBot at master • clemsonsec/skilldetective." <https://github.com/clemsonsec/skilldetective/tree/master/ChatBot> (accessed Feb. 13, 2023).

[29] Y. Wu et al., "Google's neural machine translation system: Bridging the gap between human and machine translation," arXiv Prepr. arXiv1609.08144, 2016.

## APPENDIX

TABLE VIII. POLICIES CONSIDERED BY DETBERT

Category	Policy Violation Type	Policy defined by VAs Platforms
User Privacy	<b>P1:</b> Collecting health data	Collects information related to any person's physical or mental health or condition, the provision of health care to a person, or payment for the same
	<b>P2:</b> Collecting kids' data	Collects any personal information from end users
	<b>P3:</b> Collects any sensitive personal information from end users	Collect sensitive personally identifiable information, including, passport number, social security number, national identity number, full bank account number, or full credit/debit card number
	<b>P4:</b> Lacking a privacy policy	Collect personal information from end users without providing a privacy notice that displayed in skill's detail page
Content Safety	<b>P5:</b> Toxic content	It includes content not suitable for all ages
	<b>P6:</b> Kids' safety	Promotes any content, or services, or directs end users to engage with content outside of Alexa

# Digital Stethoscope for Early Detection of Heart Disease on Phonocardiography Data

Batyrkhan Omarov<sup>1</sup>, Assyl Tuimebayev<sup>2</sup>, Rustam Abdrakhmanov<sup>3</sup>,

Bakytgul Yeskarayeva<sup>4</sup>, Daniyar Sultan<sup>5</sup>, Kanat Aidarov<sup>6</sup>

Al-Farabi Kazakh National University, Almaty, Kazakhstan<sup>1,5,6</sup>

NARXOZ University, Almaty, Kazakhstan<sup>1</sup>

INTI International University, Kuala Lumpur, Kazakhstan<sup>1</sup>

Boston University, Boston, USA<sup>2</sup>

International University of Tourism and Hospitality, Turkistan, Kazakhstan<sup>3</sup>

Khoja Akhmet Yassawi International Kazakh, Turkish University, Turkistan, Kazakhstan<sup>4</sup>

**Abstract**—The burgeoning realm of digital healthcare has unveiled a novel diagnostic instrument: a digital stethoscope tailored for the early detection of heart disease as elucidated in this research. By harnessing the nuanced capabilities of phonocardiography, this device captures intricate heart sounds, subsequently processed through advanced machine learning algorithms. Traditional stethoscopes, although indispensable, might miss subtle anomalies – a lacuna this digital counterpart addresses by meticulously analyzing phonocardiographic data for the slightest deviations indicative of cardiac anomalies. As the digital stethoscope delves into this trove of aural cues, the machine learning component discerns patterns and irregularities often imperceptible to human auditors. The confluence of these digital acoustics and computational analytics not only augments the accuracy of early heart disease diagnosis but also facilitates the archival of this data, engendering a continuous, longitudinal assessment of cardiac health. The initial foray into real-world application registered an encouraging precision rate, cementing its potential as an invaluable asset in preemptive cardiac care. With this innovation, we stand on the cusp of a paradigm shift in how heart diseases are diagnosed, making strides towards timely interventions and improved patient outcomes.

**Keywords**—Deep learning; CNN; random forest; SVM; neural network; prediction; analysis

## I. INTRODUCTION

Heart disease remains one of the foremost health challenges of the 21st century, accounting for a significant portion of morbidity and mortality rates globally [1]. Despite significant advancements in medical technology, early detection of cardiac anomalies often proves elusive, emphasizing the need for efficient, non-invasive, and accurate diagnostic tools. Traditional auscultation, using conventional stethoscopes, has been an integral part of cardiovascular assessments for nearly two centuries [2]. While these instruments have facilitated countless diagnoses, their efficacy is largely contingent on the clinician's expertise and the acoustic environment. Recognizing these limitations, there has been an increasing interest in harnessing the power of technology to augment the auditory capabilities of medical practitioners, thus making the detection process more reliable and less dependent on subjective interpretations [3].

Phonocardiography, the graphic recording of heart sounds, offers a more analytical approach to cardiac auscultation [4]. Unlike the ephemeral nature of live listening, phonocardiograms provide a tangible, visual representation of cardiac acoustics, allowing for a detailed examination of heart sound waveforms. The visual depiction of these sounds opens the door to a range of analytical possibilities, especially when combined with the vast computational power of today's digital tools [5]. However, merely converting sounds into graphs isn't sufficient for the sophisticated diagnostics required for early detection. This is where machine learning, an offspring of artificial intelligence, becomes pivotal.

Machine learning (ML) has witnessed an unprecedented surge in its applicability across various domains in the past decade [6]. In the realm of healthcare, ML algorithms are particularly valuable for pattern recognition – identifying regularities and deviations in vast datasets that would be unmanageable for humans to process manually [7]. Given the intricate nature of phonocardiographic data [8], with its myriad of subtle cues that might indicate potential pathologies, machine learning emerges as the ideal tool for deciphering this complexity. When the analytical strength of ML converges with the detailed acoustic data from a digital stethoscope, the synergy could potentially redefine the paradigms of cardiac diagnostics.

It's against this backdrop that our research ventured into developing a digital stethoscope equipped with the capacity to record phonocardiographic data, subsequently processed by state-of-the-art machine learning algorithms [9]. This innovative approach aims not only to enhance the granularity of heart sound analysis but also to democratize the diagnostic process, rendering it less reliant on individual expertise and more on objective, data-driven analytics [10]. By doing so, the intention is to unearth those elusive early markers of heart disease that, if addressed timely, could drastically alter prognostic outcomes.

In this paper, we explore the design and functionality of the digital stethoscope in question, delve into the specific machine learning algorithms employed, and evaluate the potential of this amalgamation in revolutionizing early cardiac disease detection. Through a series of trials and analyses, we aim to

underscore the instrument's diagnostic precision, its advantages over traditional auscultatory methods, and its prospective role in shaping the future of cardiac care.

## II. RELATED WORKS

The evolution of diagnostic methodologies for cardiac conditions provides a rich tapestry of innovations and paradigm shifts. Historically, the quest for early detection of heart diseases has encompassed an array of techniques, of which auscultation has been a linchpin. To understand the significance and potential impact of the digital stethoscope combined with machine learning on phonocardiography data, it's imperative to first trace the trajectory of existing literature in these domains.

### A. Traditional Auscultation and Phonocardiography: A Historical Perspective

Auscultation, the act of listening to bodily sounds, dates back to ancient times, with physicians employing rudimentary tools or direct ear placement to ascertain internal anomalies [11]. Laënnec's invention of the stethoscope in the early 19th century marked a significant leap, introducing a degree of standardization and amplification to the process [12]. Despite its ubiquity, conventional auscultation is susceptible to a range of limitations including ambient noise interference, dependence on individual auditory discernment, and the transient nature of the listening process [13].

The inception of phonocardiography sought to address some of these challenges. By providing a visual representation of heart sounds, clinicians could revisit, share, and analyze the recordings, thereby transcending the ephemerality inherent to live listening [14]. This shift to graphical cardiac sound representation allowed for a more objective and analytical approach but required adeptness in waveform interpretation [15].

### B. Digital Stethoscopes: Bridging Acoustic and Electronic Realms

As medical diagnostics progressed, so did the tools that underpin its practice. The stethoscope, a symbol of medical professionalism since the 19th century, hasn't been immune to this evolution. Its traditional acoustic counterpart, while invaluable, presented constraints in terms of sound clarity, susceptibility to ambient interference, and lacked the capability for longitudinal data recording [16].

The advent of digital stethoscopes marked a watershed moment in auscultatory practices. By incorporating electronic components, these devices promised—and often delivered—superior auditory fidelity, adeptly filtering out extraneous noises and enhancing the salience of crucial cardiac sounds [17]. Beyond mere amplification, the transformative aspect of digital stethoscopes lay in their ability to interface seamlessly with computational platforms. This not only facilitated real-time visual representation of cardiac acoustics but also opened avenues for persistent data storage, rendering sporadic health assessments a continuum of insightful cardiac monitoring [18]. While the foundational principle of listening remained unchanged, the digital shift accentuated the depth, clarity, and analytical potential of this time-honored diagnostic ritual.

### C. Machine Learning in Healthcare: A New Frontier

In the lexicon of contemporary healthcare, machine learning (ML) has rapidly ascended as a transformative force. This subset of artificial intelligence, distinguished by its capacity to autonomously evolve through data-driven insights, has opened vistas of opportunities across myriad medical domains [19].

The allure of ML in healthcare is multi-faceted. Central to its appeal is its profound capability for pattern detection, particularly salient in complex datasets where nuanced anomalies might elude human analysis [20]. Such pattern-recognition prowess has been harnessed in diverse medical terrains, from the precision of radiographic interpretations to the predictive capabilities in patient prognosis [21].

With phonocardiographic data being inherently intricate, laden with auditory subtleties indicative of potential pathologies, ML's integration in this domain has emerged as a promising frontier [22]. While the potential of machine learning is vast, its implementation in healthcare isn't merely a technological endeavor; it represents a confluence of computational excellence and clinical acumen, aspiring to reshape the contours of patient-centric care in the digital age.

### D. Integrating Machine Learning with Phonocardiography: Preliminary Endeavors

The synergy between phonocardiography and machine learning (ML) stands as an epitome of interdisciplinary convergence in modern medical research. Historically, phonocardiography, with its graphic representation of cardiac sounds, provided a tangible avenue for detailed acoustic analysis, albeit demanding meticulous human interpretation [23].

The proposition of integrating ML into this domain was fueled by the algorithmic promise of discerning intricate patterns and anomalies within these auditory datasets. Initial scholarly forays were primarily anchored in leveraging ML for extracting salient features from phonocardiographic recordings, differentiating normative heart rhythms from their pathological counterparts [24].

Subsequent research endeavors cast a wider analytical net, navigating the complexities of diverse cardiac anomalies and iterating across a spectrum of ML algorithms to optimize diagnostic accuracy [25]. Notwithstanding the promise these preliminary investigations showcased, they were often hamstrung by challenges—primarily, the quality of the phonocardiographic inputs, which were at times marred by environmental interferences or sub-optimal recording devices. Yet, these early ventures underscored the potential of this amalgamation, paving the way for the sophisticated diagnostic methodologies we envision today.

### E. Challenges and Opportunities: A Synthesis

While the confluence of digital stethoscopes and machine learning augurs well for cardiac diagnostics, challenges abound. Data privacy, especially with digitized medical records, remains a concern [26]. Additionally, ensuring algorithmic transparency and explicability in healthcare is paramount, given the stakes involved [27].

On the flip side, opportunities for this interdisciplinary venture are vast. Beyond mere diagnostics, there's potential for predictive analytics, long-term cardiac health monitoring, and even integration with telemedicine platforms, paving the way for remote diagnostics and consultations.

The literature underscores a clear trajectory: from the rudimentary act of listening to heart sounds to harnessing advanced computational tools for intricate cardiac sound analysis. The marriage of digital stethoscopes with machine learning isn't just the next step in this evolution, but potentially a giant leap, promising a future where heart disease detection is more precise, timely, and democratized.

### III. CHARACTERISTICS OF HEART SOUNDS

The cardiac sounds S1 and S2 predominantly fall within the high-frequency spectrum, optimally discerned using the diaphragm aspect of a stethoscope. A typical S1 frequency ranges from 50 to 60 Hz, while an S2 usually varies between 80 to 90 Hz [28]. On the other hand, S3 is characterized as a low-amplitude, pre-diastolic signal with a frequency band approximating 20-30Hz. S4, manifesting towards diastole's conclusion, is perceptible distinctly when utilizing a stethoscope. A deviant S4 resonates at frequencies below 20 Hz [28].

While S1 and S2 are generally detectable, their amplitude displays variability. In certain instances, due to underlying cardiac abnormalities, their audibility might be compromised. It's noteworthy that S1 and S2 do not resonate at constant frequencies but fluctuate across different cardiac cycles. These intrinsic complexities in cardiac sound demarcation have spurred scholars to architect specialized analytical methodologies [28].

Fig. 1 delineates the comprehensive categories and roles of HSs. Typically, each cardiac ailment is associated with one or two HSs. Certain anomalous heart sounds manifest as an elevated frequency noise subsequent to the primary tricuspid stenosis (TS) sound. Notably, the ejection sound (ES) is a prevalent early systolic noise, attributed to the abrupt halting of the semilunar cusps as they initiate their movement in early systole. During mid-systole, the mid-systolic click (MSC) emerges due to the abrupt cessation of prolapsing mitral valve leaflets' movement into the atrium, restrained by chordae [29].

Clinicians pay heed to these atypical cardiac sounds, recognizing their potential in providing diagnostic insights.

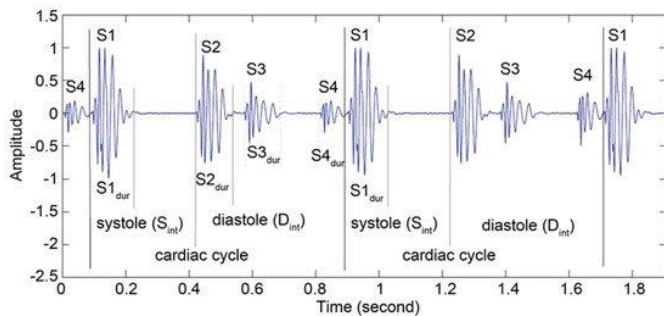


Fig. 1. Heart sounds.

### IV. ELECTRONIC STETHOSCOPE STRUCTURE

Fig. 2 presents a conceptual framework of the envisaged stethoscopic apparatus incorporating machine-learning methodologies. Heart sounds, as captured by the stethoscope, undergo amplification and filtration via an analog interface prior to their digital conversion and relay to the analytical subsystem. It's imperative for this analog interface to exhibit a superior signal-to-noise quotient, efficient common-mode suppression, and minimal baseline deviations or saturation tendencies. The pre-amplification mechanism enhances the subtle cardiac acoustic signals, initially picked up by the microphone, to a more discernible magnitude.

Fig. 3 delineates the architecture of a computer-integrated cardiac monitoring apparatus leveraging an electronic stethoscope, compartmentalized into three pivotal segments: data acquisition, pre-processing, and signal analysis. The electronic stethoscope actively records heart sounds (HS), subsequently digitized by the pre-processing segment. Within this segment, the full-frame HS signal, having undergone noise mitigation and interference reduction, is both normalized and partitioned. Signal analysis tools undertake the tasks of feature extraction and pattern categorization. The resultant structure culminates in clinically-informed diagnostic determinations. An exhaustive breakdown elucidating the intricacies and sub-components of these principal sections is provided.

#### A. Heart Sound Acquisition

Heart Sound Data Acquisition Module. The initial phase of heart sound retrieval yields automated cardiac acoustic data, serving as the foundation for subsequent processing stages.

Electronic Stethoscope Sensory Mechanism. Patient-derived cardiac acoustics are captured via a digital stethoscope, as illustrated in Fig. 4. Within this apparatus, some may employ a digital audio mechanism, a piezoelectric plate, or an aerodynamic suction module. This instrument then converts the heart's electrical impulses into auditory signals.

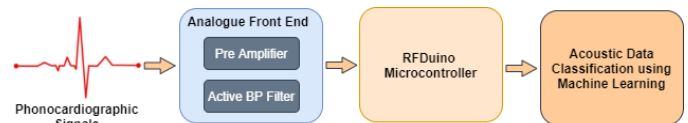


Fig. 2. Diagram of the proposed heart disease detection system.

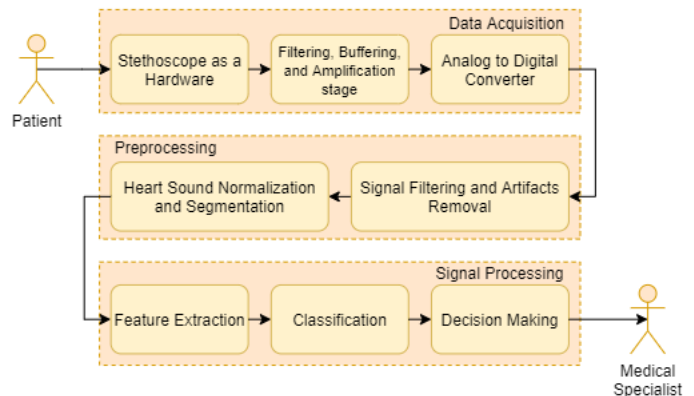


Fig. 3. Typical flow chart for heart sound signal acquisition, processing and analysis.

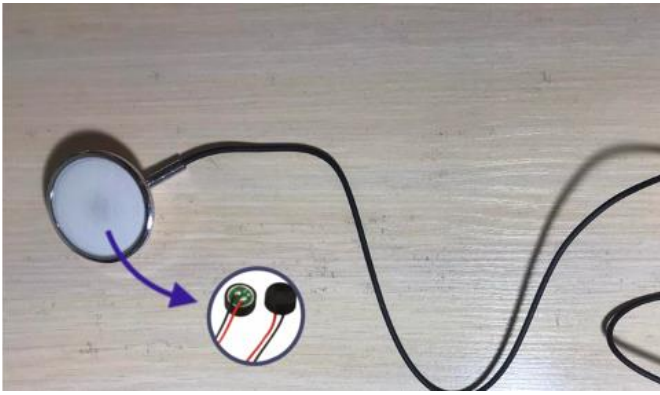


Fig. 4. Electronic sensor stethoscope.

**Amplification and Filtration Mechanism.** In diverse communication frameworks, amplification and filtration instruments are indispensable. To mitigate noise disturbances originating from power sources, a low-pass filter is employed. Subsequently, an anti-aliasing filter is integrated to reduce potential aliasing effects. Within specific system blueprints, the filtration mechanism is conceptualized as a low-pass filter, tailored to encompass the frequency spectrum of most rapid cardiac acoustics. Band-pass filtering is invoked for passband delineation, effectively countering aliasing. Post amplification, the signal undergoes digitization through an analog-to-digital transformation.

**Analog-to-Digital Transduction.** This component effectively transmutes analog signals into their digital counterparts. The parameters for this conversion can be predetermined by the equipment fabricator. Elevated bit rates and sampling frequencies can augment precision, all while economizing on bandwidth and energy consumption.

### B. Data Collection

In this stage, the digital cardiac acoustic signal undergoes reduction, standardization, and segmentation.

**Denosing Mechanism.** Typically, a digital filtration system is employed to isolate the desired signal from its embedded noise within the pertinent frequency domain. Advanced denosing methodologies are generally adopted to enhance the signal-to-noise ratio (SNR), thereby furnishing the apparatus with superior noise attenuation capabilities.

**Normalization and Cycle Division.** Diverse sampling points and processing locales often introduce variances in the captured signal during data acquisition. Consequently, cardiac sound signals undergo normalization to a predetermined scale, ensuring that data acquisition positions and multiple samples do not skew the anticipated amplitude of the signal. Post-normalization, these signals are partitioned into distinct cycles, priming them for component recognition within the heart sound and subsequent feature extraction.

### C. Heart Sound Signal Processing Module

During this phase, feature delineation and categorization activities are undertaken.

**Feature Delineation.** Signal manipulation necessitates the conversion of analog information into a digital paradigm. Such

a parametric portrayal is subsequently harnessed for in-depth analyses and applications.

**Categorization Mechanism.** After the amassed features are integrated into a classification system, it serves as a tool for data discernment, aiding healthcare practitioners in diagnostic determinations and therapeutic strategy formulations.

The processor core hubs, as depicted in Fig. 4, constitute the primary entities of an apparatus equipped to handle the digitized signal and its ensuing processing. From our meticulous investigations, it became evident that the discourse predominantly gravitates towards three pivotal phases concerning the automated identification of varied cardiac anomalies and acoustic signal afflictions: (1) Heart Sound (HS) data capture and sensory blueprinting, (2) noise attenuation and cardiac sound signal partitioning, and (3) proficient feature extraction coupled with autonomous HS analysis.

## V. PROPOSED NETWORK

For a nuanced evaluation of heart tones—encompassing rhythm, boundaries, duration, and intensity—a robust database is paramount. We curated a dataset, drawing samples primarily from heart failure patients, notably those affiliated with the Cardiology Department of the Almaty Cardiology Center. Heart tone biometric measurements were captured using an electronic stethoscope. For each individual, quintuple recordings were procured from the cardiac apex, as visualized in Fig. 5.

Ensuring database integrity is crucial for efficacious model training. Consequently, only the cardiac tones from individuals with clinically validated heart conditions were cataloged.

### A. Detection of Special Characteristics

Feature delineation serves to illuminate the distinct attributes of heart tones, facilitating differentiation of standard and anomalous cardiac sounds. An algorithm tailored for this extraction was conceived using Python, a choice made for its adeptness at signal processing functions. This algorithm, spanning eight meticulous steps, is poised to extract a rich set of attributes from a singular cardiac sound. This richness ensures individualized representation of each patient during the optimal analysis phase. It leverages a series of preset resolutions and thresholds to perform an in-depth analysis of cardiac acoustics.

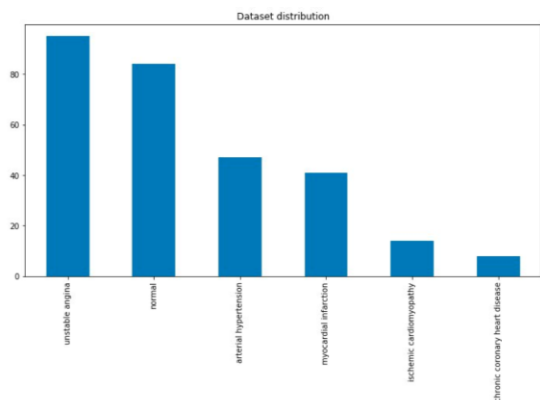


Fig. 5. Working principle of the smart stethoscope.

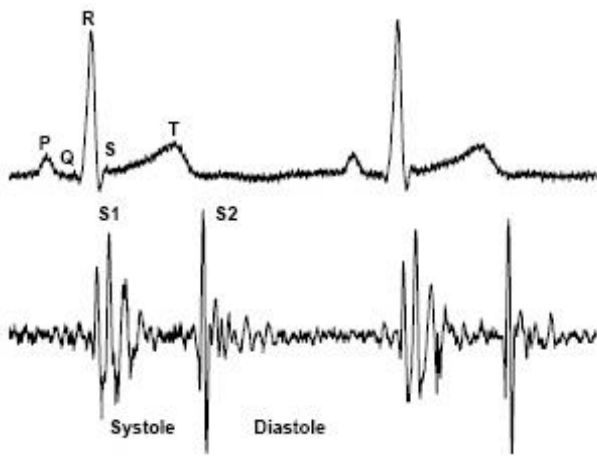


Fig. 6. Key components of cardiac sound [30].

Prevailing academic sources provide methodologies to discern the cardinal components of heart sounds, notably the primary cardiac signal (S1) and the secondary tone (S2), as well as to demarcate the boundaries of S1, S2, systole, and diastole (as depicted in Fig. 6).

In the inaugural step, a standardized root-mean-square trajectory spanning a duration of 10 seconds is derived. Given the inherently transient nature of heart sounds, 10-second segments are extracted from the primary recording to circumvent the omission of sporadic features dispersed within anomalous heart sounds. Subsequent to this, consistent RMS energy trajectories of the procured segments are crafted, primarily to accentuate the peaks of S1 and S2 whilst concurrently diminishing noise perturbations.

Step 2: Peak Recognition. The primary objective was to distinguish prominent peaks between S1 and S2 tones. Parameter resolution was employed to discern these peaks. In the event that inter-peak distances failed to align within a stipulated range, the identification process was iteratively revisited, altering the resolution to secure a consistent peak count. This iterative mechanism [31] acknowledges the variability in heart tone frequencies across individuals, correlating the resolution to each individual's unique heart rhythm. Hence, a myriad of resolutions is assessed for each cardiac sound until an optimal fit is established.

Step 3: Demarcation of Dominant Peaks. Initial steps involved assessing zero-crossings at the stipulated threshold, subsequently mapping approximate peaks for the upper boundary. If the segmented peaks did not achieve satisfactory numbers, the algorithm looped back, using a divergent resolution for prominent peak identification.

Step 4: Identification of Minute Peaks. Waves trapped between the demarcated dominant peak boundaries were analyzed, with the zenith of each wave being recognized as the minor peak. Disparities between these minor peaks and their corresponding dominant peaks were assessed. Adjustments were made if deviations exceeded acceptable ranges, and the process cyclically reverted to the dominant peak phase if necessary [32].

Step 5: Minor Peak Segmentation. Zero-crossings of the waves within the dominant peak boundaries were examined in relation to the identified minor peaks. Should any discrepancies arise in the demarcation of minor peaks, thresholds were adjusted and the segmentation was re-evaluated.

Step 6: Temporal Estimation & Validation. Consolidating prior findings, peaks were chronologically arrayed, both in terms of prominence and brevity. Rigorous validation ensured the elimination of any misaligned or overlapping peaks. In scenarios where peak counts were suboptimal, the protocol reverted to the dominant peak classification stage. This loop, targeting absolute accuracy, was exceptionally applied to cardiac tones heavily masked by noise.

Step 7: Categorization of Cardiac Tones. All segregated segments and intervals were labeled as S1, S2, systole, or diastole, resonating with the observation that systolic duration typically undercuts diastolic intervals in heart sounds.

Step 8: Feature Derivation. Post the validation process, features were distilled solely from those samples that met the requisite criteria.

### B. Applying Machine Learning for Abnormal Heartbeat Detection

Fig. 7 elucidates the amalgamation of the classification architecture alongside the signal pre-processing schematic, further incorporating machine learning methodologies. The procured heart sound data was bifurcated into datasets earmarked for model calibration and evaluation. Employing Python, both signal refinement and autonomous segmentation were executed, leading to the statistical analysis and the machine learning-driven training and categorization of cardiac tones. An overview of these pre-processing activities is delineated in the subsequent section. From the partitioned cardiac audio data, a plethora of attributes was extracted spanning time (t)-domain, frequency (f)-domain, and Mel frequency cepstral coefficients (MFCC). Prior to its introduction into the machine learning paradigm for assessment, the designated training dataset was subjected to a series of pre-processing stages.

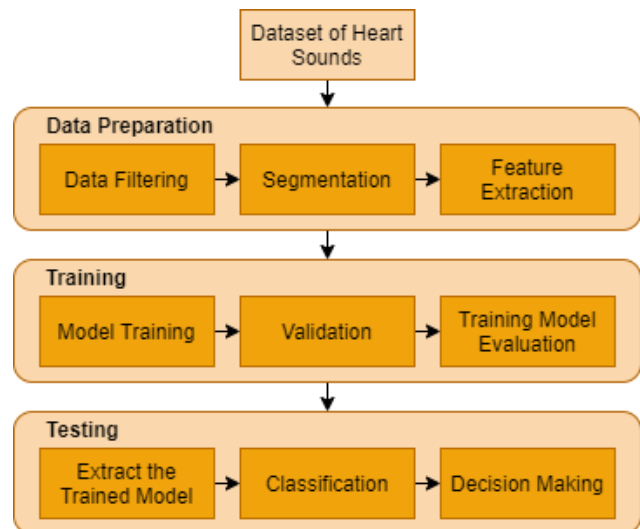


Fig. 7. Machine learning-based heart sound abnormality detection.

For the purpose of categorization, we employed the k-nearest neighbor (k-NN) classifier. Within this study, the number of neighbors and distance served as pivotal hyperparameters. Upon the meticulous extraction and mitigation of noise from the signals, we proceeded with heart sound identification. As previously highlighted, the dataset was divided, allocating 200 atypical heart sounds juxtaposed with 200 typical ones. Concurrently, the dataset was apportioned into 80% for training and 20% for testing.

The dataset bifurcation catered to two primary subsets: a training dataset and a validation dataset. The former educates the machine learning architecture utilizing samples furnished with benchmark values, aiming to curtail potential errors. In contrast, the latter evaluates the model's efficacy on previously unencountered samples, ascertaining the model's applicability to novel data. For the preliminary analysis encompassing 12 subjects, the training subset comprised 48 impulse responses, while the validation subset accounted for 46.

### C. Classification of Heart Sounds

For real-time execution, the cardiac audio data was subjected to a 10-second buffering [33-34], subsequently undergoing baseline drift rectification, segmentation into individual cardiac beats, and filtering within a specified bandwidth using Python 3.5. A multi-threaded script, penned in Python, was developed to facilitate the acquisition, buffering, real-time preprocessing, and identification of cardiac audio

data on the primary computing device. Signal preprocessing and segmentation tasks were executed on a personal computer, leveraging libraries such as Numpy, scikit-learn, and Matplotlib. The most efficacious algorithm, identified through benchmarking, was then instantiated on the personal computer for real-time classification, employing both PyBrain and Scikit-learn libraries.

## VI. EXPERIMENTAL RESULTS

### A. Hardware

The rapid advancement of mobile technologies paves the way for enhancing routine healthcare practices. Potential applications encompass leveraging mobile gadgets for clinical data collection, provisioning diagnostic information to physicians, researchers, and patients, real-time monitoring of patients' vital signs, and facilitating immediate healthcare interventions.

The designed system prioritizes minimalism, encompassing just three essential components: a stethoscope, a dedicated smartphone application, and a compact device. Within the stethoscope's chamber, an electronic microphone is strategically positioned for sound capture. To attenuate extraneous noise, all other extremities of the hose are sealed, barring the intake section. Fig. 8 delineates the constituent elements of the conceptualized stethoscope.

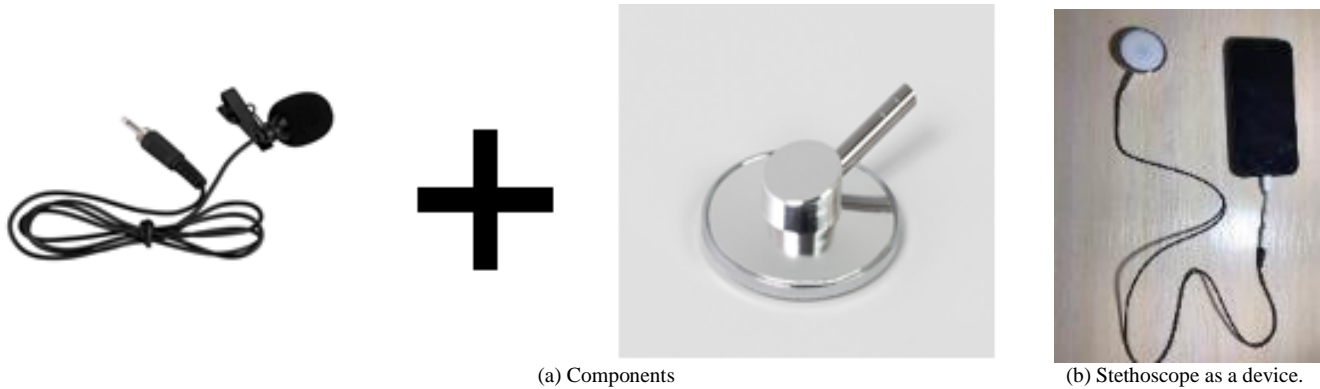


Fig. 8. Components of the proposed stethoscope.

Fig. 9 presents a comprehensive illustration of the systematic process employed for the identification of cardiac anomalies using a mobile device, following the acquisition of the heart's acoustic imprints via a stethoscope.

The initial phase entails a meticulous analysis of the audio signals captured from the stethoscope. This is succeeded by the application of a refined algorithm specifically designed to discern and neutralize extraneous ambient noise. Transitioning to the subsequent stage, a detailed classification protocol is employed to interpret these processed signals. Culminating this sequence, the analysis yields insights, upon which a potential diagnostic recommendation is formulated, providing a holistic understanding of the cardiac state.

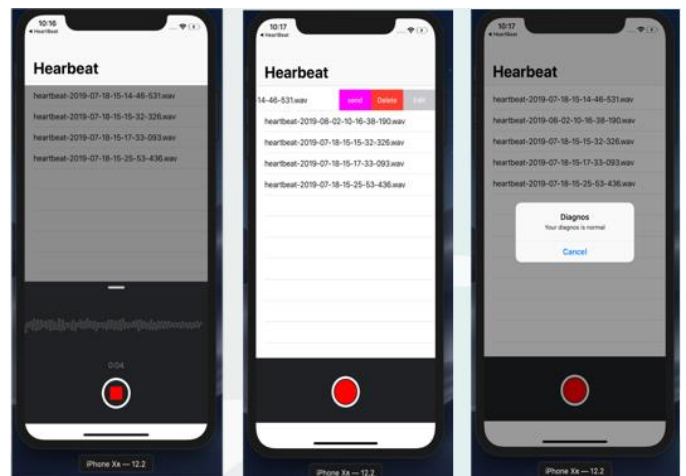


Fig. 9. Heartbeat abnormality detection process.

B. Classification Results

Fig. 10 depicts various cardiac acoustic patterns. These include: Normal: Typical sounds indicative of a healthy heart. Murmur: Additional auditory phenomena resulting from perturbations in blood flow, producing discernible vibrations. Extrahls: An ancillary auditory signature. Artifacts: A diverse array of distinct auditory emissions.

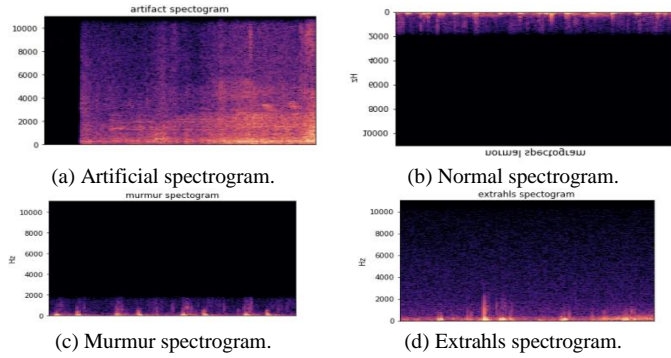


Fig. 10. Time domain PCG trace and its power spectral density for different types of heart sounds.

Fig. 11 depicts the training and validation processes utilized for the identification of atypical heart rhythms. This representation provides insights into both training and validation accuracy metrics over a span of 300 epochs. In Fig. 12, the evolution of training and validation loss metrics across the training epochs is showcased. Notably, post approximately 100 epochs, both the training and validation losses appear to stabilize.

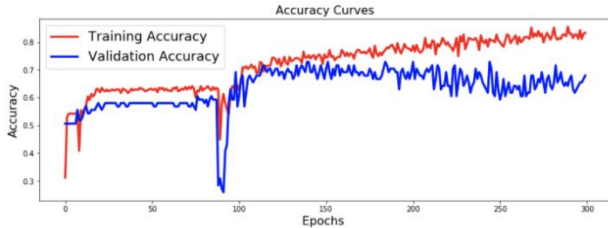


Fig. 11. Model training and validation for abnormal heartbeat detection.

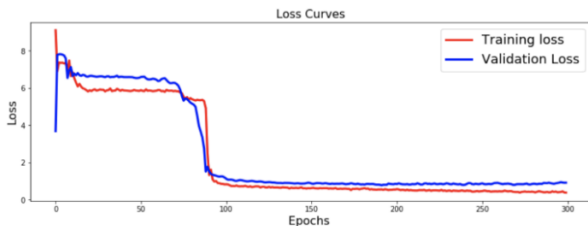


Fig. 12. Model training and validation for abnormal heartbeat detection.

Fig. 13 presents a confusion matrix detailing the classification outcomes for five distinct cardiac conditions: normal heartbeat, murmur, extrasystole, extrahls, and artifacts. The findings underscore a commendable level of precision in classifying cardiac sound patterns and pinpointing abnormal heart rhythms.

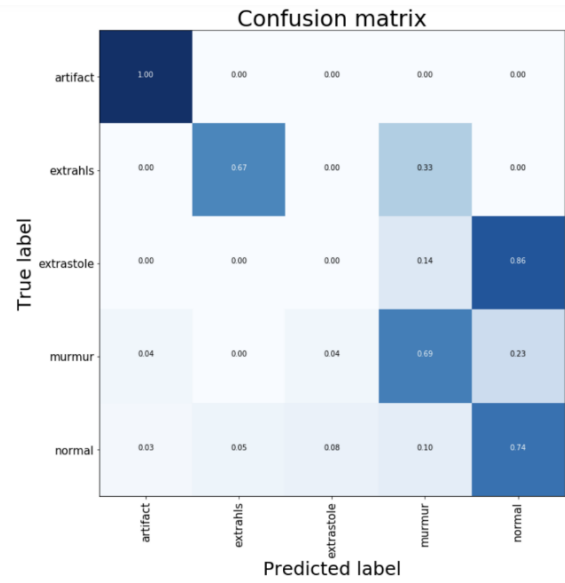


Fig. 13. Confusion matrix.

VII. DISCUSSION

The examination of heart sounds, especially as an early diagnostic tool, has been a significant area of interest in cardiological research. This study's central tenet sought to harness the technological capabilities of today's world, combined with the intricacies of cardiac acoustics, to devise a model that could reliably identify and classify abnormal heart rhythms.

One of the most notable findings of this study was the efficacy with which the model was able to delineate between different heart conditions. With the increasing prevalence of cardiovascular diseases globally, having an accessible and precise diagnostic tool can potentially revolutionize cardiac care, especially in areas where specialized cardiac care remains elusive. The use of smartphones, as indicated in this research, points towards a trend in telemedicine and mobile health (mHealth) solutions, which have gained substantial traction over recent years.

The study's multi-step approach, starting from data collection using an electronic stethoscope to signal preprocessing, feature extraction, and final classification, ensured a comprehensive review of the heart sounds. The detailed steps, as represented in the figures, allow for a meticulous understanding of how the model refines and uses the data. This approach is essential, especially given the critical nature of the data in question; heart sounds are not only diverse but also nuanced.

One of the primary challenges faced in many similar studies is the noise interference in heart sound recordings. Our methodology, which incorporated sophisticated noise reduction techniques, was able to significantly improve the signal-to-noise ratio (SNR). The pre-processing of heart sound signals, as delineated in our study, presents a robust method of ensuring the integrity of the data, further enhancing the model's reliability.



The utilization of Python, a versatile and widely adopted programming language, underscores the scalability and adaptability of the proposed method. The integration of multiple libraries like Numpy, scikit-learn, and Matplotlib not only facilitated rigorous data processing but also ensures that the model can be integrated or adapted into various other platforms or studies.

The training, testing, and validation datasets' demarcation ensures that the model is not just accurate but also generalizable. Often, machine learning models might overfit to the training data, making them less reliable when exposed to new, unseen data. Our model, after undergoing rigorous training and validation, demonstrated a commendable degree of accuracy, pointing towards its robustness.

However, it is also essential to note the limitations of this study. The presented model, while advanced, is primarily dependent on the quality of the initial recordings. Factors like the positioning of the stethoscope, the ambient environment, and the patient's physical condition can introduce variances in the recorded sounds. Furthermore, while the study encapsulated multiple heart conditions, there remains a wide variety of cardiac anomalies, each with its unique acoustic signature, which might not be entirely accounted for in this research.

Additionally, while smartphones and mobile applications promise a more democratized healthcare landscape, their efficacy is inherently tied to factors like smartphone penetration in a region, digital literacy, and the reliability of digital infrastructures. Hence, while the model offers promise, its large-scale application would require a more ecosystem-driven approach, ensuring that all potential bottlenecks are addressed.

In conclusion, the presented research underscores the potential of merging technology and cardiology, offering a glimpse into the future of cardiac diagnostics. The methodology, marked by its rigor and attention to detail, sets a precedent for further studies in this domain. Future research might look into integrating more varied heart sound datasets, exploring the potential of real-time diagnostics, and even combining this acoustic data with other diagnostic metrics for a more comprehensive assessment. The horizon of cardiac care, augmented by technology, seems promising, and this research serves as a beacon in that journey.

## VIII. CONCLUSION

This research undertook the ambitious endeavor of bridging the realms of advanced technological tools with the intricate field of cardiology, underscoring the transformative potential such intersections hold for modern medicine. The primary focus was the identification and classification of heart sounds, tapping into the ever-evolving capabilities of machine learning and the widespread accessibility of smartphones.

The developed model, as showcased in this study, demonstrated notable accuracy in deciphering and distinguishing between various heart conditions. These findings are of paramount importance, especially considering the global rise in cardiovascular diseases and the resultant need for accessible, accurate, and timely diagnostic tools. The utility

of a smartphone-based diagnostic mechanism extends beyond mere convenience; it potentially democratizes cardiac care, paving the way for early interventions even in areas bereft of specialized healthcare infrastructures.

However, it is essential to recognize that while the results are promising, the journey is only just beginning. The marriage of technology and healthcare, though filled with potential, also demands a rigorously holistic approach. Factors ranging from the quality of data acquisition to the challenges associated with the mass adoption of smartphone-based medical tools must be addressed for this research's broader implications to fully materialize.

In summary, this study has laid down a robust foundation, emphasizing the confluence of technology and cardiology as a potent avenue for future research and applications. As we look ahead, it becomes evident that the future of cardiac care, supported by technological innovations, has the potential to reshape healthcare landscapes, making diagnostics more accurate, accessible, and timely. This research stands as a testament to that potential, signaling an exciting trajectory for both cardiac care and medical technology.

## REFERENCES

- [1] M. E. Chowdhury, A. Khandakar, K. Alzoubi, S. Mansoor, A. Tahir et al., "Real-time smart-digital stethoscope system for heart diseases monitoring," *Sensors*, vol. 19, no. 12, pp. 2781, 2019.
- [2] M. Elhilali and J.E. West, "The stethoscope gets smart: Engineers from Johns Hopkins are giving the humble stethoscope an AI upgrade," *IEEE Spectrum*, vol. 56, no. 2, pp. 36-41, 2019.
- [3] M.N. Türker, Y.C. Çağan, B. Yildirim, M. Demirel, A. Özmen et al., "Smart Stethoscope," In 2020 Medical Technologies Congress, pp. 1-4, 2020.
- [4] Y.J. Lin, C.W. Chuang, C.Y. Yen, S.H. Huang, P.W. Huang et al., "An intelligent stethoscope with ECG and heart sound synchronous display" In 2019 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 1-4, 2019.
- [5] Ş. Tekin, "Is Big Data the new stethoscope? Perils of digital phenotyping to address mental illness," *Philosophy & Technology*, pp. 1-15, 2020.
- [6] V.T. Tran and W.H. Tsai, "Stethoscope-sensed speech and breath-sounds for person identification with sparse training data," *IEEE Sensors Journal*, vol. 20, no. 2, pp. 848-859, 2019.
- [7] 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.
- [8] H. Bello, B. Zhou and P. Lukowicz, "Facial muscle activity recognition with reconfigurable differential stethoscope-microphones," *Sensors*, vol. 20, no. 17, pp. 4904, 2020.
- [9] Sultan, D., Omarov, B., Kozhamkulova, Z., Kazbekova, G., Alimzhanova, L., Dautbayeva, A., ... & Abdrakhmanov, R. (2023). A Review of Machine Learning Techniques in Cyberbullying Detection. *Computers, Materials & Continua*, 74(3).
- [10] K.A. Babu and B. Ramkumar, "Automatic detection and classification of systolic and diastolic profiles of PCG corrupted due to limitations of electronic stethoscope recording," *IEEE Sensors Journal*, 2020.
- [11] V. Arora, R. Leekha, R. Singh and I. Chana, "Heart sound classification using machine learning and phonocardiogram," *Modern Physics Letters B*, vol. 33, no. 26, pp. 1950321, 2019.
- [12] S. Vernekar, S. Nair, D. Vijaysenan and R. Ranjan, R, "A novel approach for classification of normal/abnormal phonocardiogram recordings using temporal signal analysis and machine learning," In 2016 Computing in Cardiology Conference (CinC), pp. 1141-1144, 2016.

- [13] M.N. Homsy and P. Warrick, "Ensemble methods with outliers for phonocardiogram classification," *Physiological measurement*, vol. 38, no. 8, pp. 1631, 2017.
- [14] G. Son and S. Kwon, "Classification of heart sound signal using multiple features," *Applied Sciences*, vol. 8, no. 12, pp. 2344, 2018.
- [15] M. Chowdhury, A. Khandakar, K. Alzoubi, S. Mansoor, A. Tahir et al., "Real-time smart-digital stethoscope system for heart diseases monitoring," *Sensors*, vol. 19, no. 12, pp. 2781, 2019.
- [16] M. Suboh, M. Yaakop, M. Ali, M. Mashor, A. Saad et al., "Portable heart valve disease screening device using electronic stethoscope," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 15, no. 1, pp. 122-132, 2019.
- [17] V. Varghees and K. Ramachandran, "Effective heart sound segmentation and murmur classification using empirical wavelet transform and instantaneous phase for electronic stethoscope," *IEEE Sensors Journal*, vol. 17, no. 12, pp. 3861-3872, 2017.
- [18] J. Roy, T. Roy and S. Mukhopadhyay, "Heart sound: Detection and analytical approach towards diseases," *Smart Sensors, Measurement and Instrumentation*, vol. 29, no. 1, pp. 103-145, 2019.
- [19] 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.
- [20] A. Alqudah, H. Alquran and I. Qasmieh, "Classification of heart sound short records using bispectrum analysis approach images and deep learning," *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 9, no. 1, pp. 1-16, 2020.
- [21] S. Singh, T. Meitei and S. Majumder, "Short PCG classification based on deep learning," In *Deep Learning Techniques for Biomedical and Health Informatics*, pp. 141-164, 2020.
- [22] Y. Khalifa, J. Coyle and E. Sejdic, "Non-invasive identification of swallows via deep learning in high resolution cervical auscultation recordings," *Scientific Reports*, vol. 10, no. 1, pp. 1-13, 2020.
- [23] H. Li, X. Wang, C. Liu, Q. Zeng, Y. Zheng et al., "A fusion framework based on multi-domain features and deep learning features of phonocardiogram for coronary artery disease detection," *Computers in biology and medicine*, vol. 120, pp. 103733, 2020.
- [24] Baikuvekov, M., Tolep, A., Sultan, D., Kassymova, D., Kuntunova, L., & Aidarov, K. (2023). 1D Convolutional Neural Network for Detecting Heart Diseases using Phonocardiograms. *International Journal of Advanced Computer Science and Applications*, 14(3).
- [25] Ahmad, Z., Zeeshan, M., Sohail, A., Haris, M., Khan, M. U., Hussain, S. S., & Khan, M. S. (2023, February). Automatic Detection of Paediatric Congenital Heart Diseases from Phonocardiogram Signals. In *2023 3rd International Conference on Artificial Intelligence (ICAI)* (pp. 188-195). IEEE.
- [26] Rong, Y., Fynn, M., Nordholm, S., Siaw, S., & Dwivedi, G. (2023, July). Wearable Electro-Phonocardiography Device for Cardiovascular Disease Monitoring. In *2023 IEEE Statistical Signal Processing Workshop (SSP)* (pp. 413-417). IEEE.
- [27] Roy, T. S., Roy, J. K., & Mandal, N. (2023). Design and development of electronic stethoscope for early screening of valvular heart disease prediction. *Biomedical Signal Processing and Control*, 86, 105086.
- [28] F.D.L. Hedayioglu, "Heart Sound Segmentation for Digital Stethoscope Integration," Master's Thesis, University of Porto, Porto, Portugal, 2011.
- [29] H. Li, G. Ren, X. Yu, D. Wang and S. Wu, "Discrimination of the Diastolic Murmurs in Coronary Heart Disease and in Valvular Disease," *IEEE Access*, pp. 160407-160413, 2020.
- [30] A. Yadav, A. Singh, M.K. Dutta and C.M. Travieso, "Machine learning-based classification of cardiac diseases from PCG recorded heart sounds," *Neural Computing and Applications*, pp. 1-14, 2019.
- [31] R. Banerjee, S. Biswas, S. Banerjee, A.D. Choudhury, and T. Chattopadhyay et al., "Time-frequency analysis of phonocardiogram for classifying heart disease," In *2016 Computing in Cardiology Conference (CinC)*, pp. 573-576, 2016.
- [32] L.G. Durand and P. Pibarot, "Most recent advancements in digital signal processing of the phonocardiogram," *Critical Reviews™ in Biomedical Engineering*, vol. 45, pp. 1-6, 2017.
- [33] Omarov, B., Narynov, S., Zhumanov, Z., Gumar, A., & Khassanova, M. (2022). State-of-the-art violence detection techniques in video surveillance security systems: a systematic review. *PeerJ Computer Science*, 8, e920.
- [34] Omarov, B., Tursynova, A., Postolache, O., Gamry, K., Batyrbekov, A., Aldeshov, S., ... & Shiyapov, K. (2022). Modified UNet Model for Brain Stroke Lesion Segmentation on Computed Tomography Images. *Computers, Materials & Continua*, 71(3).

# Predicting the Level of Safety Feeling of Bangladeshi Internet users using Data Mining and Machine Learning

Md. Safiul Alam, Anirban Roy, Partha Protim Majumder, Sharun Akter Khushbu

Department of Computer Science and Engineering, Daffodil International University, Dhaka, Bangladesh

**Abstract**—An amazing combination of cutting-edge data mining and machine learning methodologies to predict the level of safety feeling among Bangladeshi internet users, which is a significant departure in this subject. By leveraging cutting-edge algorithms and innovative data sources, this work provides previously unheard-of insights into how this demographic perceives online safety, shedding light on an essential yet underappreciated aspect of their digital lives. This exceptional study's original research increases the body of knowledge of online safety and sets the road for policy recommendations and intervention tactics that will enable Bangladesh to become a global leader in internet security.

**Keywords**—Bangladesh; data analysis; data mining; important factors; machine learning; prediction; performance evaluation metrics; safety level

## I. INTRODUCTION

Every day, more people are using the internet than ever before, all over the world [1]. This rate is increasing in Bangladesh too [2]. Because recently Bangladesh has seen growth in internet usage [3]. At present, the Internet has become a massive part of people's daily lives in Bangladesh [4]. As a result, communication, business, education, banking, service, jobs, etc. are turning online day by day in Bangladesh [5]. In March 2021 Internet users in Bangladesh increased to 116 million whereas the population of Bangladesh at the time was 167 million which means 70% of the population had access to the internet [6]. People want to feel safe, secure, and devoid of any bullying, harassment, and illegal activity when using the internet [7]. According to a UNICEF survey, 32% of Bangladeshi children, aged 10 to 17, are familiar with and have encountered online abuse, harassment, and cyberbullying. 25% of them have access to the internet by the age of 11. Additionally, according to a Telenor Group and Grameenphone report, online bullying is a serious problem for 85% of Bangladeshi youngsters. According to the report, 18% of them experienced worse bullying as a result of the shocking COVID-19 epidemic [8]. So, it is very important to know people's safety feelings at the time of using the internet [9]. The advent of the digital era has created possibilities and challenges never before experienced, altering how people connect, communicate, and access information throughout the globe. As the internet continues to permeate every area of our daily lives, online safety and security have become a huge concern [10]. For that it is essential for the user to feel safe while using the internet [11]. There is a great depiction of an

accurate prediction of an individual's safety level at the time of using the internet is indispensable with prior knowledge about the important factors, which have a great impact [12]. Moreover, it is necessary for private and public organizations, industries, banks, and IT companies to find out people's safety level at the time of using the Internet [13]. Because it will make their services more secure and effective [14]. In that case, safety level prediction will act as a guide to making an appropriate safety level which has been fulfilled in this research.

This work closes a huge knowledge gap that has mostly gone unfilled up to this point. Despite the abundance of research on online safety, there are surprisingly few that focus on the unique viewpoints and experiences of internet users in Bangladesh [15]. Because of its geographical emphasis and dedication to illuminating the intricacies of online safety in the context of Bangladesh, this study is a pioneering effort that stands out [16]. This work's significance extends beyond the sphere of academic research; it has a significant impact on Bangladesh's evolving digital ecosystem and tackles urgent problems that have not yet been fully investigated [17]. This groundbreaking study demonstrates its importance in a number of ways. In the age of digital transformation, where internet access is nearly widespread, it is imperative to provide Bangladeshi internet users with the knowledge and tools they need to properly navigate the online world [18]. By predicting people's safety feelings and fostering a sense of control and confidence in the face of online hazards, this study strengthens people's agency in safeguarding their digital experiences [19]. For Bangladeshi internet service providers, regulators, and legislators, the novel approach of this study offers a once-in-a-lifetime chance to tailor safety measures and actions [20]. Knowing the specific factors influencing safety feelings may help them develop more effective strategies and policies that match the local context, which will eventually lead to a safer online environment. The integration of data mining and machine learning in this study has increased the prevalence of data-driven decision-making in the area of internet safety. The efficacy of organizations and authorities may be improved by using the information gathered from this study to assist them in allocating resources and choosing initiatives based on actual evidence. The lack of study on the perspectives of Bangladeshi internet users on online safety highlights the novelty and significance of this endeavor [21]. This study investigates an understudied area, filling a large gap in the literature and setting the stage for future studies that have an

emphasis on regional and local variability. As Bangladesh embraces digitization, developing a culture of cybersecurity becomes increasingly important [22]. This work has the potential to promote best practices among internet users, academic institutions, and businesses by igniting dialogues on online safety. Also, this study combines data mining and machine learning, fusing cutting-edge technology with real-world applications. The innovative approaches adopted might serve as a paradigm for future studies on the intersection of data science and cybersecurity in Bangladesh and elsewhere in the world [23]. Even though this study's findings are anchored in the context of Bangladesh, they may still be applicable to other developing nations that are going through rapid digitalization [24]. This research's importance transcends national boundaries since the technique and results developed here may be changed and applied in similar situations. There isn't a single, universal approach to online safety [25]. This unique piece of work is actually innovative since it allows for the customization of safety precautions. By anticipating Bangladeshi users' safety attitudes and allowing interventions and assistance to be personalized to individuals' particular concerns and experiences, online safety is made more pertinent and effective. It also reveals the attitudes and beliefs of Bangladeshi internet users, shedding light on a hitherto unresearched facet of online safety. This highlights the emotional and intangible aspects of cybersecurity that are occasionally overshadowed by technology solutions [26]. By detecting and evaluating these emotions, this approach improves our understanding of the human side of cyber security. The originality of this work opens the door for future research initiatives that focus on the feelings and experiences of internet users in a variety of contexts. It sets a precedent for appreciating the importance of the human element in cybersecurity and might ignite a larger conversation about the psychological aspects of online safety [27]. Along with being creative, it may help the Bangladeshi online community understand online safety by making it more pertinent, relatable, and human.

Additionally, this approach combines the strengths of machine learning and data mining. By exploiting the capabilities of this cutting-edge technology, the research proposes a creative way of predicting safety feelings that are tailored to the Bangladeshi environment. Combining these methods should result in conclusions that are more accurate, and practical, and represent a novel contribution to the field of internet safety.

Data mining and machine learning, a subfield of artificial intelligence (AI), employ statistical techniques to give computers the capacity to learn from data and improve their performance on certain tasks [28]. Data mining and machine learning are used to enable learning and inference across a heterogeneous mix of devices, including PCs, smartphones, IoT devices, and edge devices [29]. A data mining and machine learning probabilistic system is a complex tool that may be used to evaluate obtained data, provide predictions or judgments based on that data, and then present those findings to the user [30].

As Bangladesh continues its journey toward digital transformation, the findings of this study have the potential to

inform governmental decisions, empower internet service providers to enhance user safety, and ultimately create a more secure online environment. By bridging the gap between data-driven insights and the specific problems faced by Bangladeshi internet users, this research provides a groundbreaking contribution to maintaining the online experiences of an expanding online community. Here, emphasis has been given to the analysis of some empirical factors of an individual's data to perform the safety level prediction.

In this research, safety level predictions have been done and several factors behind this have been analyzed. Moreover, extensive research and analysis have been conducted. Here, several data mining techniques have been applied for experimentation, and several performance evaluation metrics to evaluate this work. Twelve popular data mining classifiers, including Logistic Regression, MLP, KNN, Decision Tree, Naive Bayes, Search Vector Machine, Gradient Boosting, Linear Discriminant Analysis, Stochastic Gradient Descent, Ada Boosting, Bagging, and Random Forest, have been experimented with on a survey dataset. Several performance evaluation metrics have been calculated to determine the best classifier in the working context, and a result comparison is presented here. From the analysis of the obtained result, it is confirmed that the Decision Tree classifier achieves the best result in terms of metrics.

These are the order of this paper: Section II gives an exhaustive overview of relevant studies. The study methodology is presented in Section III along with a brief overview of the dataset, data analysis, implementation process, classifiers, the outcome of the experiment, and additional findings while the conclusion is given in Section IV. Finally, Section V provides future work.

## II. LITERATURE REVIEW

The ultimate purpose of this research work is to the safety level of the user. After going through several articles, it is discovered that there has been no existing work like this done before. However, it has been unable to locate a compass in the large ocean of scholarly works that might direct us through the uncharted area of understanding how Bangladeshi internet users view their online safety [31]. This absence emphasizes the originality and importance of this research, which aims to address this important gap and improve not only the scholarly community but also the daily lives of countless Bangladeshi internet users [32]. The awareness that addressing the safety concerns of internet users goes beyond academic study and constitutes a necessary first step in establishing a more secure and safe online environment for everyone has steered this research down an innovative route [33]. To implement this unique model, some papers have been studied. All of them are described below as per the research paper's theme:

Syeda et al. [34] applied seven approaches of data mining i.e. KNN, Decision Tree, SVM, NN, Naive Bayes, Logistic Regression, and Random Forest to predict user satisfaction and dissatisfaction. They have taken different parameters which are produced with high accuracy. The accuracy for KNN, Decision Tree, SVM, NN, Naive Bayes, Logistic Regression and, Random Forest were 96%, 93.33%, 93.3%,

86%, 89.3% and, 96%. Though the highest accuracy is achieved by three algorithms, they have chosen Random Forest because it shows better precision, recall, and f1 score rather than others.

In order to identify phishing websites, Kaytan et al. [35] suggested a clever model based on extreme learning machines. Website forms differ from one another in terms of how they perform. Therefore, they must make use of special web page features to prevent phishing assaults. Additionally, they proposed a template based on computer training methods for identifying phishing web pages. The model has one output and 30 inputs. In this application, the 10-fold cross-validation test was run. The classification's total accuracy was 95.05 percent.

Salehin et al. [36] Karim advocated combining LSTM and artificial intelligence to produce a straightforward rainfall forecast model. The correctness of the deep learning approach is essential for this manner of application has been established. They used 6 parameters in their article. The accuracy was increased to 76% by looking at all the data integrating LSTM and artificial intelligence to produce a straightforward rainfall forecast model. They used 6 parameters in their article. The accuracy was increased to 76% by looking at all the data.

Salehin et al. [37] recommended utilizing RHMCD as a model to assist machine learning algorithms accomplishes the intended goal. Naive Bayes classifiers, logistic regression, and support vector machines are the algorithms that were tested. The sentiment analysis method was employed to gather information on mental health issues. The amount of depression was assessed using the decision tree method.

Salehin et al. [38] predicted the severity of depression caused by excessive cell phone use. Depression is detected using the Linear Regression technique and two machine learning algorithms, decision trees.

Technologies for agriculture have been created by Salehin et al. [39]. Various viral, fungal, and bacterial illnesses result in a significant loss of agricultural produce. In this research, they categorize crop situations based on various datasets by using the Scale Invariant Transform Feature (SIFT) technique. Finally, the solution was made available through live online portals and SMS services.

Talha et al. [40] draw attention to the significant drawback and its many root causes, including emotional instability, despair, stress, and loneliness. Physical, virtual, and medical reports were the three approaches that were used to collect the data. The detrimental impact of human behavior is demonstrated by the 71% optimistic theorem of Naive Bayes. For measurement purposes in search vector machine (SVM), negative and positive parameters are set. Last but not least, they compare the outcomes of our suggested specialization to those of the three fundamental points of reference.

Syeda et al. [41] used educational data mining to forecast the pupils' success. The entirety of the projection was based on the students' current location and general academic standing.

Yeasin et al. [42] suggested using the data mining approach to forecast students' careers. Only CS grads have had this task completed for them. They used a number of classifiers, and the accuracy varied according on each classifier. Just 506 data records were used in this study, and no distinct training or testing datasets were indicated.

Alonzo et al. [43] provided a thorough analysis of how different machine learning algorithms are used to predict and rate the quality of coconut sugar.

Perez et al. [44] provided examples of the findings from a case study on educational data analytics that was focused on identifying undergraduate students majoring in systems engineering who had dropped out after six years of attendance. Their experimental findings demonstrated that straightforward algorithms may identify dropout predictors with sustained levels of accuracy. Here, the output of four algorithms—decision trees, logistic regression, naive bayes, and random forest—was examined to suggest the best course of action. The major findings are presented here to lower the dropout rate by identifying probable causes. In addition, they provided some evaluations of the data's quality to help the students refine their data collection techniques.

With the purpose of resolving the dropout prediction problem, Mi et al. [45] developed different temporal models. Specifically, based on substantial research conducted with a few massive open online courses (MOOCs) accessible through edX and Coursera. They claimed that one logical improvement to the model, which would include a max pooling layer before the output layer, would further their work. They anticipated that their model's expansion would increase its robustness.

Aksenova et al. [46] reported an enrollment prediction research that uses support vector machines and rule-based predictive models with the aim of predicting the overall enrollment headcount, which is made up of continuing, returning, and new (freshman and transfer) students. The core prediction findings are generated using a machine learning approach called SVM, which is then applied by a program called Cubist to create simple rule-based predictive models. Lastly, they provided some experimental findings about the forecasting of student enrolment.

### III. METHODOLOGY

This section is parted into Data Description, Data Collection, Data Preprocessing, Data Analysis, Classifier Description, Implementation Procedure, Result and Discussion, and Evaluation. This section basically presents the approach taken to accomplish this work.

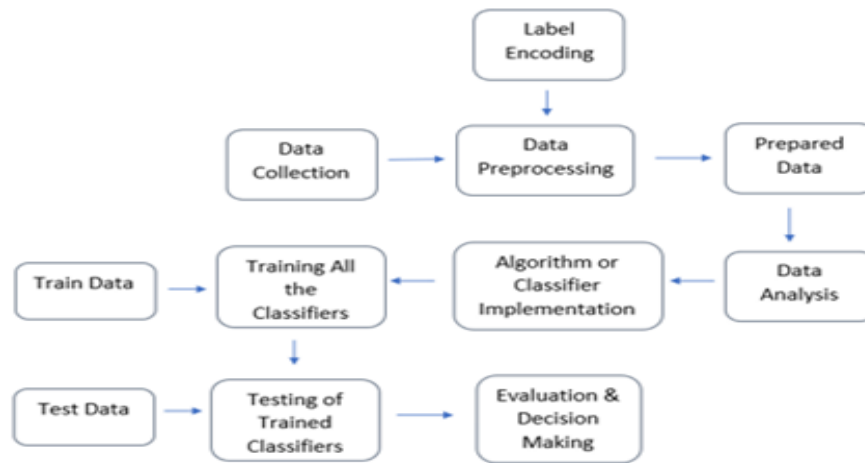


Fig. 1. Methodology diagram.

For this work, several steps have been performed, as presented in Fig. 1. A detailed description of all the subsections is presented below.

### A. Data Description

Information that approximates and characterizes is referred to as qualitative data. It is possible to notice and document qualitative data [47]. In statistics, qualitative data is sometimes referred to as categorical data since it can be categorized based on the characteristics and traits of an object or phenomenon [48]. Any information that can be quantified and employed in statistical or mathematical calculations is referred to as quantitative data [49]. Making judgments in real life using mathematical inferences is aided by this type of data [50]. So, in this work, all the data are qualitative before preprocessing, and after preprocessing, they are converted to quantitative data for analysis and to build a machine learning model for prediction. A decision has been made after evaluation.

### B. Data Collection

Data is survey-based data. The survey has been performed. Most of the data has been collected by physical survey and some of the data has been collected through an online survey. A total of 5,321 individual records are used here to accomplish this work. The survey mainly consists of 8 questions.

### C. Data Preprocessing

After checking for null values, it has been found that there have been no missing values in the dataset as all the answers to the 8 questions have been obtained from the respondents and the information has been carefully compiled to make the dataset. Fig. 2 shows that there are no missing values. The data type information has been checked, and it has been observed that 6 columns have object-type values. The label encoding pre-processing technique has been used to convert these object-type values into numeric. Among the 8 questions, 7 questions



Fig. 2. Heatmap for checking null values.

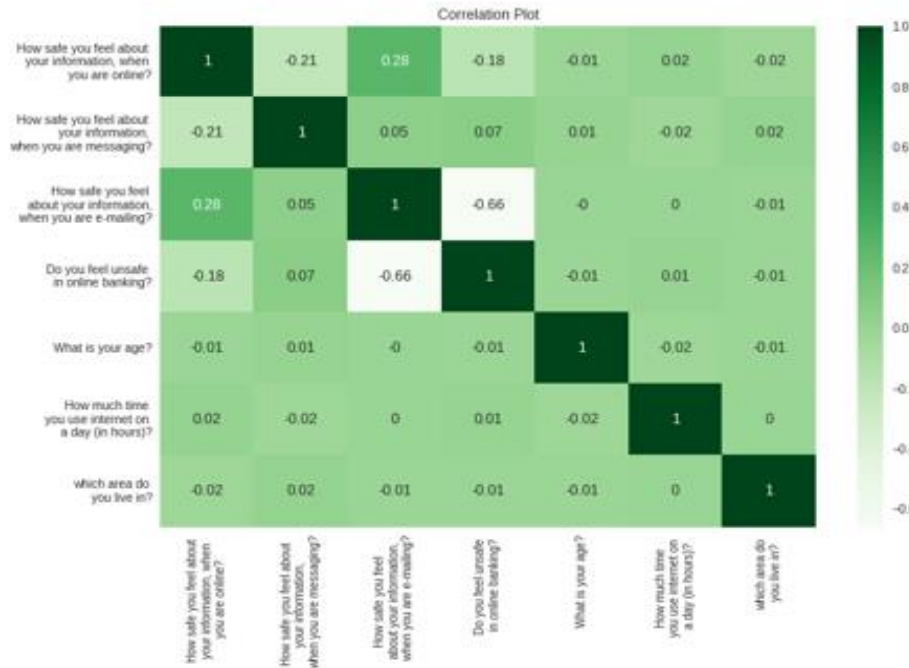


Fig. 3. Correlation matrix.

(How safe you feel about your information, when you are online? How safe you feel about your information, when you are messaging? How safe you feel about your information, when you are e-mailing? Do you feel unsafe in online banking? What is your age? How much time you use internet on a day (in hours)? which area do you live in?) This has been used as the independent variable and only one question (Safety\_Level) has been used as the dependent variable.

To prevent overfitting, the dataset has been split into training and testing sets. The correlation of the independent variables in the training dataset has then been determined, as shown in Fig. 3. A total of 73% of the data has been used for the training of the classifier and 27% has been employed for testing purposes. To retrieve appropriate attributes, a threshold value of 0.78 has been set. Using this value, it has surprisingly been found that the 8 features that have been used as the independent variables do not need to be changed.

D. Data Analysis

Data analysis is the process of cleansing, converting, and modeling data to discover useful information for commercial decision-making. The goal of data analysis is to gather useful data and make decisions based on that analysis. When determining what is occurred most recently in real life or how something plays out when making a certain decision, a simple illustration is provided of how the data is interpreted.

In a survey of 5,321 respondents, it is discovered that 31.20% of individuals feel extremely safe about their information when they are online, 32.14% of people feel no safety at all, and 36.65% of people feel poor safety about it. These results are depicted in Fig. 4.

Fig. 5 shows the results of a survey of 5,321 people, which is revealed that 35.38% of respondents feel only moderately safe about their information when messaging, 32.33% of respondents feel no safety at all, and 32.29% of respondents feel extremely safe about their information when messaging.

Fig. 6 illustrates the results of a survey of 5,321 respondents, which is surprisingly revealed that 48.74% of them feel only somewhat secure sending information through e-mail, 26.82% feel no safety at all, and 24.44% feel extremely safe sending information via e-mail.

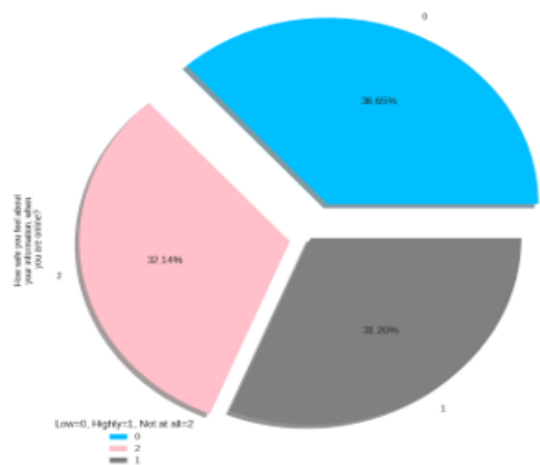


Fig. 4. People’s safety feeling about their information while using the internet.

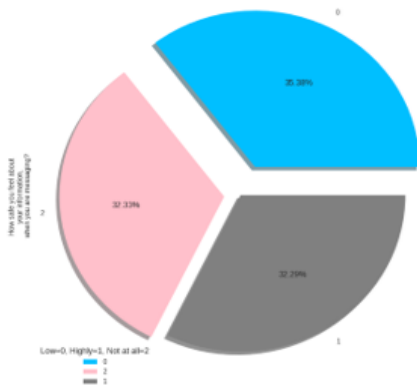


Fig. 5. Safety feeling while messaging.

The results depicted in Fig. 7 demonstrate that 57.50% of the 5,321 respondents feel unsafe while using internet banking, while the remaining 42.50% feel secure.

The bar in Fig. 8 shows the number of observations for each of the five potential category value combinations. It can be observed that individuals who feel less secure about their information while online are given a lower Safety Level rating than those who feel more secure and those who feel absolutely no security at all. Additionally, it is found that individuals who do not feel secure about their information when online seldom perceive their Safety Level to be high, while those who feel extremely secure about their personal data while online have rated their Safety Level as higher than low.

The bar in Fig. 9 displays the number of observations for each of the five potential category value combinations. It can be observed from the figure that individuals who feel the least safe when texting are assessed to have a lower Safety Level than those who feel the safest and those who feel the least safe while messaging. Additionally, it is surprising that people seldom perceive their Safety Level to be as high when they are texting using the internet when they do not feel safe and feel uncomfortable about their information when they are texting. However, those who feel extremely secure about the privacy of their information when communicating have rated their Safety Level as the highest.

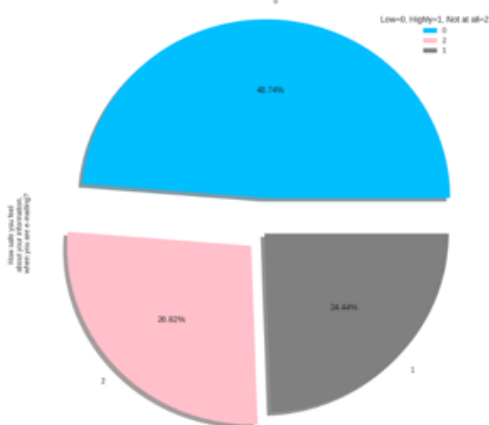


Fig. 6. Sense of safety while e-mailing.

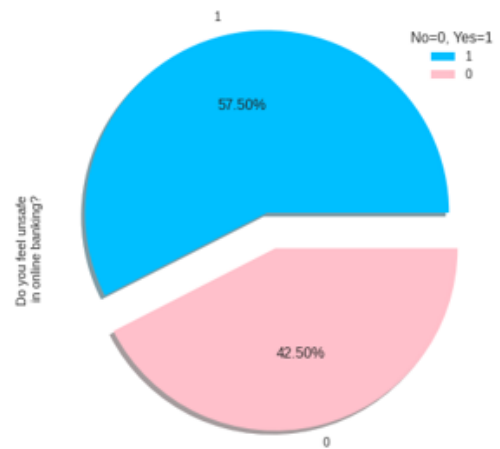


Fig. 7. People's thinking of online banking.

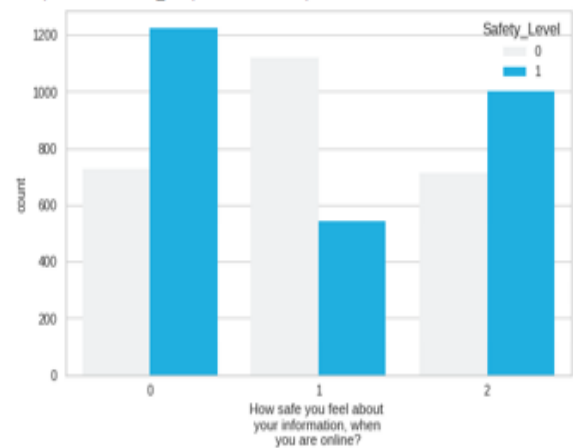


Fig. 8. Impact of First Attribute on Target Variable.

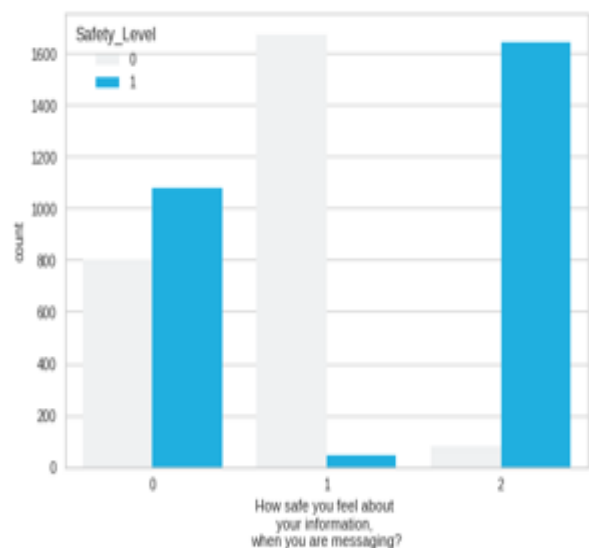


Fig. 9. Second Attribute's Effect on the Target Variable.



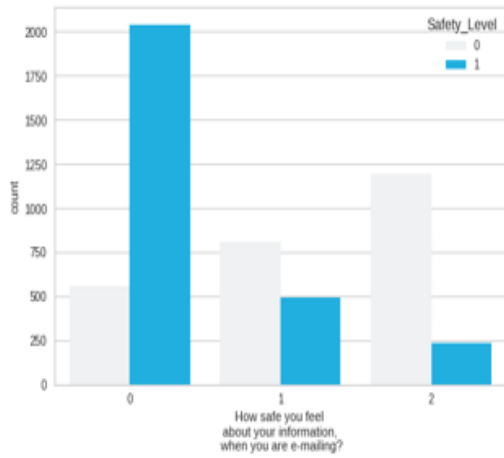


Fig. 10. Influence of the third attribute on the target variable.

Fig. 10 displays the counts of observations for each of the five potential category value combinations. It can be seen that individuals who feel less safe about their information when emailing have a lower Safety\_Level rating than those who feel more secure and those who feel no security at all. Moreover, people who feel less safe about their information when emailing rarely perceive their Safety\_Level to be high. In contrast, it has found that people who feel unsafe about their information when emailing consider their Safety\_Level to be the highest.

The bar chart in Fig. 11 shows the number of observations for each of the four potential category value combinations. It can be observed that individuals who feel unsafe while conducting online banking rated their Safety\_Level lower than those who feel secure. Interestingly, individuals who feel unsafe when using online banking rarely rated their Safety\_Level as the highest. However, those who feel secure when using internet banking have rated their Safety\_Level as the highest.

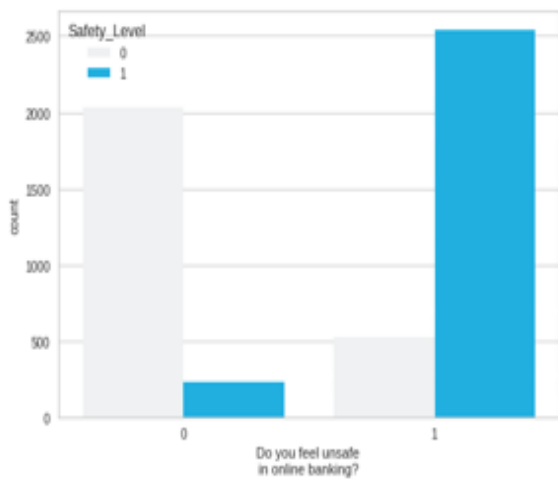


Fig. 11. Significance of the fourth attribute on the target variable.

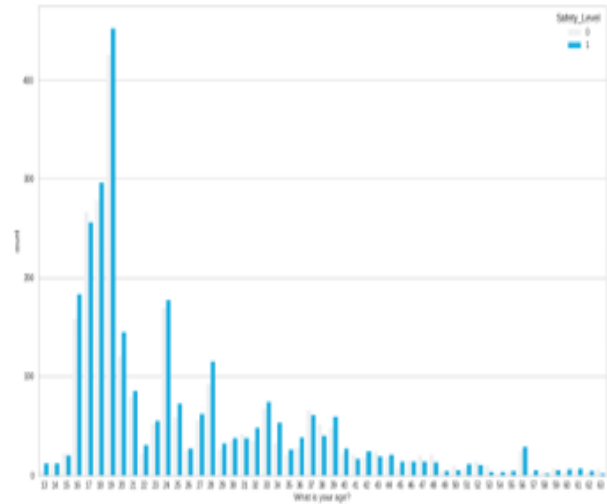


Fig. 12. Role of the sixth attribute on the target variable.

Fig. 12 demonstrates that individuals in the following age groups have believed their safety level to be low: 13 to 14, 16, 18 to 30, 32 to 36, 39 to 40, 42, 44 to 45, 51, 54, 56 to 57, and 60 to 62. On the other hand, those between the ages of 15, 17, 31, 37 to 38, 41 to 43, 46 to 48, 50 to 52, 53 to 58, and 63 thought their safety level is high. Interestingly, respondents between the ages of 49 and 56 are perceived their safety level to be both high and low.

Fig. 13 shows that the people who spend 3 to 4 hours, 7 to 13 hours, and 16 hours a day using the internet have considered their Safety\_Level as low. On the other hand, people who spend 2 hours, 5 to 6 hours, and 14 hours a day using the internet have considered their Safety\_Level as high.

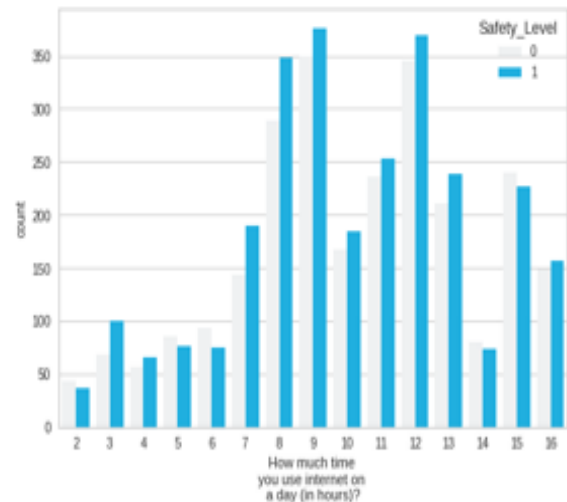


Fig. 13. Importance of the fifth attribute on the target variable.

Among 5,321 respondents, it is found that 51.99% of people consider their Safety\_Level as low while they are using the internet and 48.01% of people consider their Safety\_Level as high, as shown in Fig. 14.

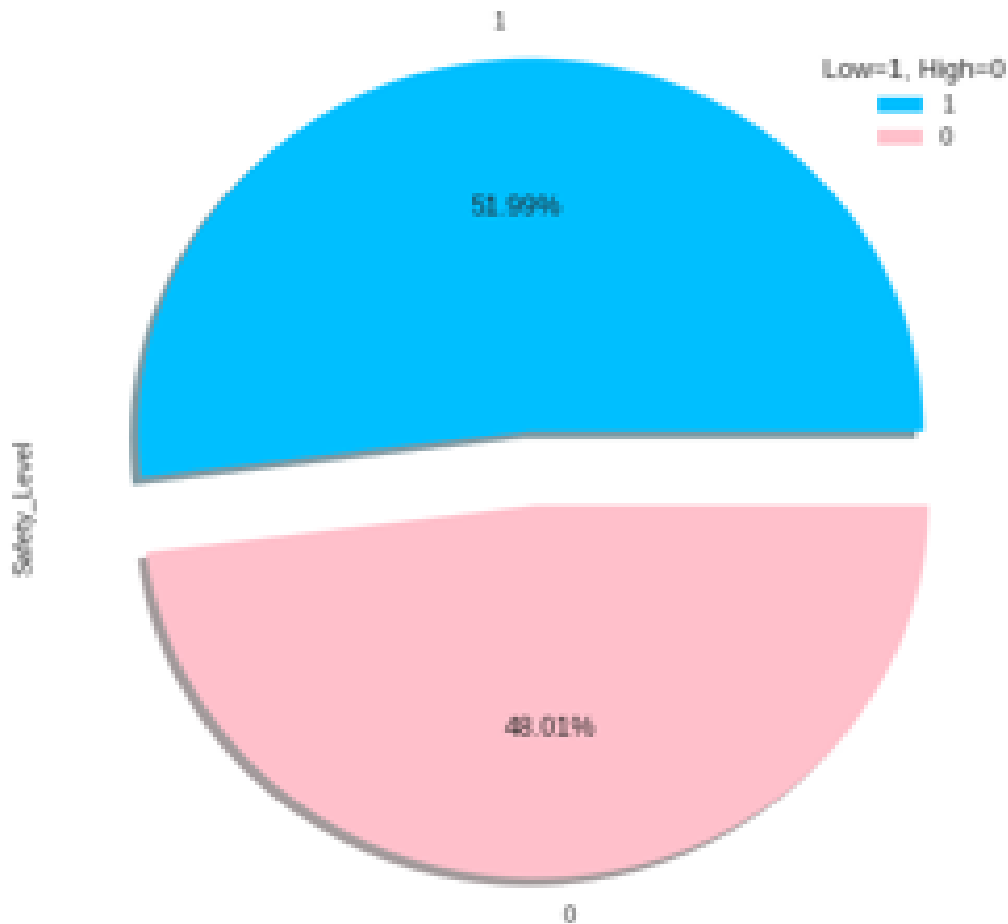


Fig. 14. People's consideration of safety level while using the internet.

### E. Classifier Description

A classifier in machine learning is a tool for forecasting the target characteristic from feature data points. Twelve classifiers have been used to analyze the dataset, and the following theory is pertinent.

In this work, Naive Bayes classifier has been used. This classifier employs Bayesian classification methodologies. It applies Bayes' theorem to class prediction and determines the class-conditional probability by taking into account the fact that the attributes are conditionally independent given the class label. This classifier can handle binary and multiclass classification issues since predictors all make independent assumptions. The work primarily focuses on a binary classification problem.

A Multilayer Perceptron (MLP) consists of an input, an output, and a number of hidden layers (one or more). Single-layer perceptrons can only learn linear functions, but multi-layer perceptrons can learn nonlinear functions. The MLP learning procedure is known as the Backpropagation Algorithm. Once the input layer receives the signal, the output layer anticipates making a decision based on the input [51]. The hidden layers serve as the computational engine for estimating continuous functions [52]. The output of one layer in an MLP serves as the input for the layer that follows.

The optimal division for each node is selected using local knowledge via a greedy method known as a Decision Tree. One conclusion is that a better tree may be produced by altering the divisional components [53]. It is well known that trees are incredibly flexible and exhibit little distortion in their interactions.

The decision tree classifier was designed particularly for the ensemble method known as Random Forest. The main function of the random forest classifier is to integrate the predictions of several trees (decision trees), where each decision tree is constructed from the output of a different dataset of random vectors. Problems with grouping are generally resolved with it. Using data samples, Random Forest algorithms build decision trees, predict those trees, and then let users vote on the best result. The group technique is superior to a single tree since it supports the outcome and reduces over-adjustment.

A statistical approach for analyzing a data set containing one or more independent variables that affect the outcome is logistic regression. To assess the result, a dichotomous variable is employed in this (only two possible outcomes). The goal of this classifier is to choose the model that best depicts, using the logistic function as support, the connection between the outcome variable and the predictor factors.

To address two-group classification issues, supervised machine learning models called support vector machines (SVM) use classification methods. Once provided with a set of labeled training data for each category, an SVM model can classify incoming text. They perform better with fewer samples and are more effective, which are their two main advantages (in the thousands). The method works well for text classification problems since it is customary to only have access to datasets with a small number of tags on each sample.

The k-nearest neighbors method, often known as KNN, is a supervised learning classifier that makes predictions or classifications about how a single data point will be grouped. It is frequently used as a categorization strategy since it is predicated on the notion that similar points could be found adjacent to one another. The k parameter of the k-NN algorithm determines how many neighbors will be looked at in order to categorize a certain query point. If k=1, for example, the case will be put in the identical class as it's only nearest neighbor.

Bagging, often referred to as bootstrap aggression, is an effective collective tactic. A technique for combining the results of different machine-learning algorithms to create predictions that are more accurate is called an ensemble approach. A broad method known as bootstrap aggregation may be used to minimize variation in algorithms with a lot of it. As with hybrid approaches like classification and regression, bagging has a large variance. A high-variance machine learning system, like decision trees, is exposed to the Bootstrap technique during the bagging process.

A quick and effective method for training linear classifiers and regressors under convex loss functions is stochastic gradient descent (SGD). SGD has been present in the machine learning field for a while, but in the context of large-scale learning, it has just lately attracted a lot of interest. Because the update to the coefficients is done for each training instance rather than at the end of examples, it has been successfully used for large-scale datasets. The Stochastic Gradient Descent (SGD) classifier essentially implements a straightforward SGD learning method that supports multiple classification loss functions and penalties.

In 1996, Freund and Schapire proposed AdaBoost. By transforming a number of weak learners into strong learners, these methods increase prediction ability. It creates a classifier by combining a number of subpar classifiers. Each iteration involves training the data and setting the classifier weights.

The combination of gradient descent and boost is known as Gradient Boosting. Each new model in gradient boosting employs the gradient descent method to reduce the loss function from its forerunner. This process is repeated until the target variable's estimation becomes even better. In contrast to previous ensemble approaches, gradient boosting builds a succession of trees, each one attempting to fix the flaws of the one before it.

For supervised classification issues, a dimensionality reduction method called Linear Discriminant Analysis is frequently employed. It is used to represent group distinctions, i.e. to distinguish between two or more classes [54]. In a lower

dimension space, it is used to project the characteristics from a higher-dimension space. In order to save money and dimensions, this can be used to project characteristics from higher dimensional space into lower dimensional space.

#### F. Implementation Procedure

The aims of this work are to perform the safety level prediction and to analyze the important factors behind choosing a particular safety level for an individual. Many significant parameters are considered here to ensure an effective prediction.

The work primarily focuses on a binary classification problem. A questionnaire form containing 8 questions was created and data was collected from different professions of people and many random people through this questionnaire. Preprocessing techniques were used to feed this data into the classifier. To label the answer to the particular question, numbers (e.g. 0, 1) were used. The dataset had a variable/attribute named "Safety\_Level" with two possible outcomes High (0) and Low (1). After preprocessing, the prepared data was partitioned into the training and testing set. 73% of the data from the total dataset was used for training purposes and the rest 27% of the whole dataset was used for testing purposes. The classifiers were trained with the training data and then used to predict the Safety\_Level using both the test data and train data. Metrics were calculated for the performance evaluation and the best classifier was determined based on the confusion matrix generated by the classifier.

#### G. Result and Discussion

In this section, the experimental result and the discussion of the obtained result of the study are presented. The result of the confusion matrix for the test data of twelve classifiers is tabulated in Table I. Since it is a two-class problem, so the classifiers generate a 2\*2 matrix.

At the time of implementation, 1,437 respondent instances are put into the testing set where the actual safety level of 667 students is high or positive. On the other hand, the actual safety level of 760 respondents is low or negative. After implementation, it has been found that a confusion matrix for each classifier which is stated in Table I. The experimental result of the confusion matrix in detail for the most competent classifier and the worst classifier has been found from Table I. From Table I, it has been found that the decision tree classifier is correctly able to predict that 607 respondents will be considered their safety level as high among 667 respondents. So, the rest of the 70 respondents among the 667 respondents are incorrectly classified that they will not be considered their safety level as high. On the other hand, this classifier is correctly able to predict that 729 respondents will be considered their safety level as low among 760 respondents. So, the rest of the 31 respondents among the 760 respondents are incorrectly classified that they do not be considered their safety level as low. From Table I, it has been found that the decision tree classifier is correctly able to predict that 530 respondents will be considered their safety level as high among 667 respondents. So, the rest of the 147 respondents among the 667 respondents are incorrectly classified that they will not be considered their safety level as high. On the other hand, this classifier is correctly able to predict that 578

respondents will be considered their safety level as low among 760 respondents.

So, the rest of the 182 respondents among the 760 respondents are incorrectly classified that they do not be considered their safety level as low. From Table I it has been found that the MLP algorithm has the highest specificity which is 0.98. On the other hand, the KNN algorithm has the lowest specificity which is 0.76. Specificity means the true negative rate. In this work, the specificity of a classifier refers to how well a classifier identifies respondents who will be considered their safety level as low. Decision Tree has 0.96 specificity means that it can identify 96% of respondents consider their safety level as low. From the value of the

confusion matrix, a classification report, macro average, and weighted average of test data for each of the classifiers has been computed which are presented in Table II and Table III. From Table II it has found that the precision of the MLP classifier for the High class is 0.97 and of the Bagging classifier for the Low class is 0.93 which are the highest, the recall of the MLP classifier for the Low class is 0.98, and of the Bagging classifier for High class is 0.92 which are the highest, and the f1-score of the Decision tree classifier for High and Low class is 0.92, 0.94 which are the highest. From Table III, it has surprisingly found that the Decision Tree classifier has the highest precision, recall, and f1-score. On the other hand, the KNN classifier has the lowest precision, recall, and f1-score.

TABLE I. CONFUSION MATRIX AND SPECIFICITY RESULT OF THE TWELVE WORKING CLASSIFIER

Classifier Name	True Positive	False Negative	False Positive	True Negative	Specificity
Decision Tree	607	70	31	729	0.96
Random Forest	606	71	54	706	0.93
Naive Bayes	562	115	84	676	0.89
Logistic Regression	590	87	89	671	0.88
KNN	530	147	182	578	0.76
SVM	594	83	68	692	0.91
Gradient Boosting	602	75	40	720	0.95
Stochastic Gradient Descent	574	103	73	687	0.90
Linear Discriminant Analysis	587	90	89	671	0.88
MLP	568	109	17	743	0.98
Ada Boost	601	76	43	717	0.94
Bagging	628	49	92	668	0.88

TABLE II. CLASSIFICATION REPORT OF ALL THE TWELVE CLASSIFIERS

Classifier Name	Class Name	Precision	Recall	F1-Score	Support
Decision Tree	Low	0.91	0.96	0.94	760
	High	0.95	0.90	0.92	677
Random Forest	Low	0.91	0.93	0.92	760
	High	0.92	0.90	0.91	677
Naive Bayes	Low	0.86	0.89	0.87	760
	High	0.87	0.83	0.85	677
Logistic Regression	Low	0.89	0.88	0.88	760
	High	0.87	0.87	0.87	677
KNN	Low	0.80	0.76	0.78	760
	High	0.74	0.78	0.76	677
SVM	Low	0.89	0.91	0.90	760
	High	0.90	0.88	0.89	677
Gradient Boosting	Low	0.91	0.95	0.93	760
	High	0.94	0.90	0.91	677
Stochastic Gradient Descent	Low	0.87	0.90	0.89	760
	High	0.89	0.85	0.87	677
Linear Discriminant Analysis	Low	0.89	0.88	0.88	760
	High	0.87	0.87	0.87	677
MLP	Low	0.87	0.98	0.92	760
	High	0.97	0.84	0.90	677
Ada Boost	Low	0.90	0.94	0.92	760
	High	0.93	0.89	0.91	677
Bagging	Low	0.93	0.88	0.91	760
	High	0.87	0.92	0.90	677

TABLE III. MACRO AVERAGE AND WEIGHTED AVERAGE OF ALL THE TWELVE CLASSIFIERS

Classifier Name	Macro Average				Weighted Average			
	Precision	Recall	F1- Score	Support	Precision	Recall	F1- Score	Support
Decision Tree	0.93	0.93	0.93	1437	0.93	0.93	0.93	1437
Random Forest	0.91	0.91	0.91	1437	0.91	0.91	0.91	1437
Naive Bayes	0.86	0.86	0.86	1437	0.86	0.86	0.86	1437
Logistic Regression	0.77	0.77	0.77	1437	0.77	0.77	0.77	1437
KNN	0.92	0.92	0.92	1437	0.92	0.92	0.92	1437
SVM	0.88	0.88	0.88	1437	0.88	0.88	0.88	1437
Gradient Boosting	0.88	0.88	0.88	1437	0.88	0.88	0.88	1437
Stochastic Gradient Descent	0.92	0.91	0.91	1437	0.92	0.91	0.91	1437
Linear Discriminant Analysis	0.88	0.88	0.88	1437	0.88	0.88	0.88	1437
MLP	0.92	0.91	0.91	1437	0.92	0.91	0.91	1437
Ada Boost	0.92	0.92	0.92	1437	0.92	0.92	0.92	1437
Bagging	0.90	0.90	0.90	1437	0.90	0.90	0.90	1437

TABLE IV. AUROC SCORE OF TWELVE CLASSIFIERS

Classifier Name	AUROC Score
Decision Tree	0.983
Random Forest	0.914
Naive Bayes	0.913
Logistic Regression	0.950
KNN	0.846
SVM	0.949
Gradient Boosting	0.977
Stochastic Gradient Descent	0.887
Linear Discriminant Analysis	0.942
MLP	0.981
Ada Boost	0.962
Bagging	0.940

TABLE V. USED PARAMETERS AND ACCURACY OF TWELVE CLASSIFIERS

Classifier Name	Parameter Detail	Accuracy (For Test Data)	Accuracy (For Train Data)
Decision Tree	max_depth=6	0.93	0.93
Random Forest	n_estimators=1	0.91	0.96
Naive Bayes	alpha=1.0, fit_prior=True	0.86	0.86
Logistic Regression	random_state=1	0.88	0.88
KNN	n_neighbors=3	0.77	0.89
SVM	probability=True, kernel='linear'	0.89	0.89
Gradient Boosting	n_estimators=88, learning_rate=1.0,max_depth=1, random_state=0	0.92	0.93
Stochastic Gradient Descent	loss="modified_huber"	0.88	0.88
Linear Discriminant Analysis	n_components=1	0.88	0.88
MLP	random_state=1, max_iter=300	0.91	0.88
Ada Boost	n_estimators=105	0.92	0.92
Bagging	n_estimators=2, random_state=0	0.90	0.96

Table IV shows us the AUROC score for each classifier. AUC means the area under the curve which helps to understand the performance of the model [55]. From Table IV it has been found that the Decision Tree classifier has the highest AUROC score which is 0.983. On the other hand, KNN has the lowest AUROC score which is 0.846.

Table V represents the accuracy of all algorithms for both training data and testing data. Also, Table V illustrates the

parameters and the different things that are used in this work to implement the algorithms selected. These parameters have been taken for better accuracy. After analyzing Table V in other words after comparing the accuracy of test data and train data for each classifier it can ensure that there is no overfitting and underfitting situation in this model [56]. The highest accuracy for test data is 0.93 which is achieved by Decision Tree. On the other hand, the lowest accuracy for test data is 0.77 which is achieved by KNN.

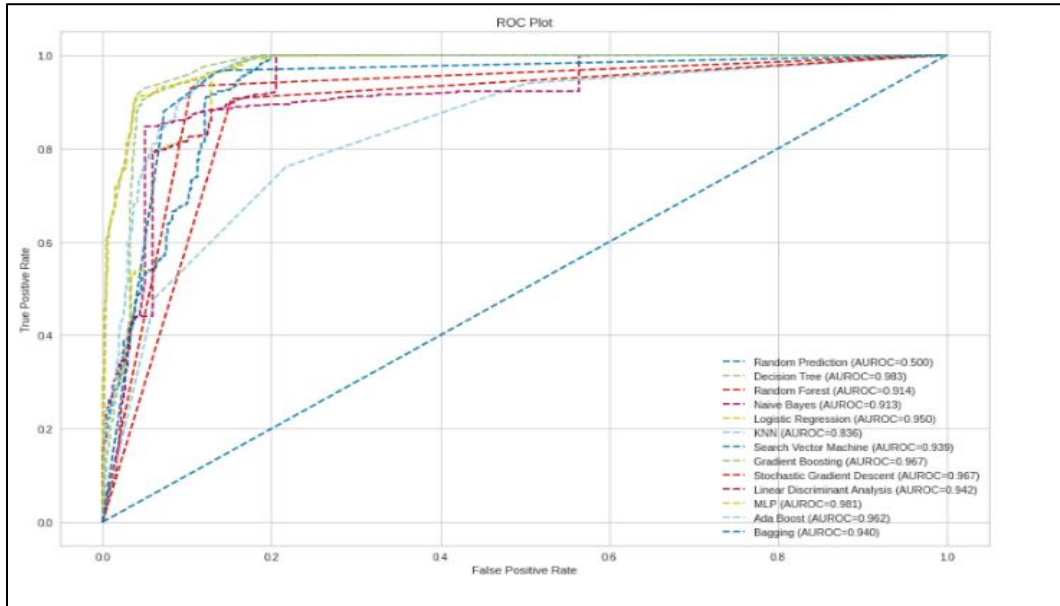


Fig. 15. ROC graph of all the twelve classifiers.

Fig. 15 shows the ROC. ROC means receiver operating characteristic which has been helped to evaluate the performance of diagnostic tests [57]. The blue line actually cuts diagonally across the rectangle here across a call which is actually a random classification that is made not based on any classifier so it simply splits the data into two so it is based on chance [58]. Also, in the blue line, the recall and specificity are equal. Fig. 15 has been made from Table IV where it has been seen that the Decision Tree classifier gives the highest performance than others. It has also been found that the KNN classifier gives the lowest performance than any other classifier.

Table VI provides a list of each algorithm’s name, the mean accuracy, and the standard deviation accuracy. From the above table, it has amazingly found that four algorithms that have the highest mean accuracy for train data which are Decision Tree, Gradient Boosting, MLP, and Ada Boost. These four algorithms’ mean accuracy is 0.92. On the other hand, the KNN classifier has the lowest mean accuracy for train data which is 0.77. From the above table, it has also been found that the Stochastic Gradient Descent is the highest standard deviation accuracy for train data which is 0.08. Also, it has surprisingly found that Decision Tree, Gradient Boosting, MLP, and Ada Boost have the lowest standard deviation accuracy for train data which is 0.01.

Fig. 16 shows the comparison of different algorithms which have been used to build the model. From these results,

it is suggested that Decision Tree, Gradient Boosting, MLP, and Ada Boost are perhaps worthy of further study on this problem.

TABLE VI. MEAN ACCURACY AND STANDARD DEVIATION OF TWELVE ALGORITHMS

Algorithm Name	Mean Accuracy (For train data)	Standard Deviation Accuracy (For Train Data)
Decision Tree	0.92	0.01
Random Forest	0.89	0.02
Naive Bayes	0.85	0.02
Logistic Regression	0.87	0.02
KNN	0.77	0.03
SVM	0.86	0.02
Gradient Boosting	0.92	0.01
Stochastic Gradient Descent	0.81	0.08
Linear Discriminant Analysis	0.87	0.02
MLP	0.92	0.01
Ada Boost	0.92	0.01
Bagging	0.90	0.02

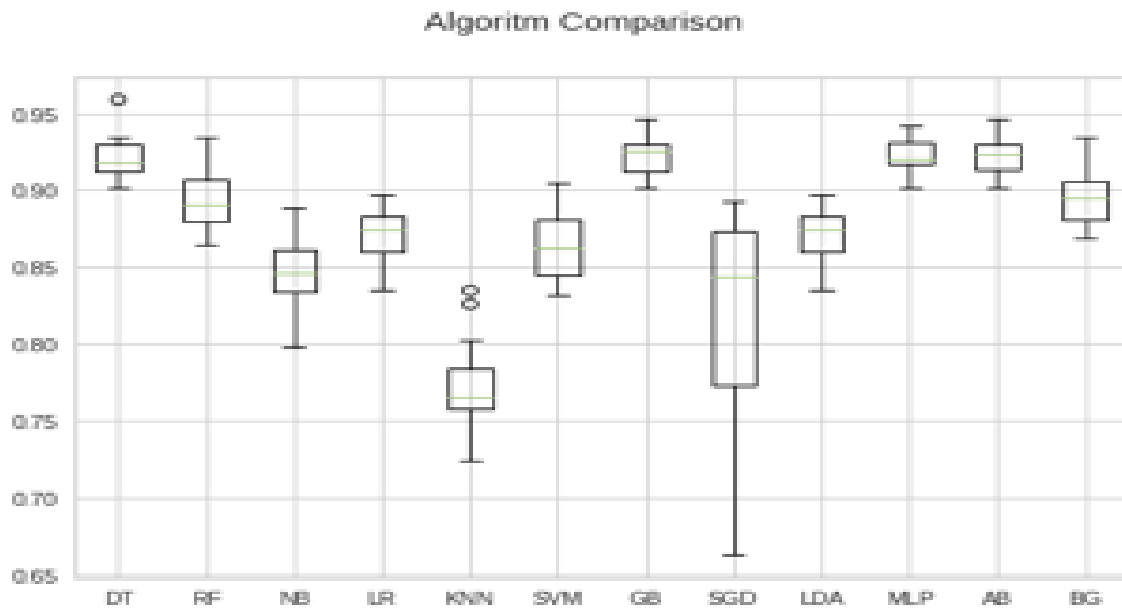


Fig. 16. Comparing all the twelve classifiers by using boxplot.

#### H. Evaluation

Comparison of training accuracy and testing accuracy is very important to understand the overfitting situation and underfitting situation in a machine learning model [59]. However, most of the previous research works had not shown the comparison of the test accuracy and train accuracy of their model which has been the main reason for being unable to verify their model's performance properly. This problem has been solved in this amazing piece of work and has been shown in Section III (G). The Decision Tree algorithm achieved the highest accuracy of 0.93. Also, based on the results analyzed in Section III (G), this algorithm was chosen as the final algorithm.

#### IV. CONCLUSION

The major goals of this work are to anticipate a person's level of online safety feeling and to identify the deciding elements that affect that person's decision to select a specific level of internet safety feeling. It is concluded from the analysis of the collected data that 48.01% of individuals feel extremely safe while using the internet, compared to 51.99% who don't, which raises serious concerns for the future growth of the nation. A variety of data mining approaches are used. A total of 73% and 27% of the data are used to train and test the classifier, respectively, in order to complete this task. A number of performance assessment measures are examined to gauge how well the functional classifier performed. The decision tree classifier surpasses conventional data mining algorithms.

#### V. FUTURE WORK

It is speculated that Decision Tree, Gradient Boosting, MLP, and Ada Boost are probably worthy of additional investigation on this subject based on Fig. 16 in section III (G).

#### ACKNOWLEDGMENT

Gratitude is expressed to each and every responder who assisted with data gathering by sharing information. Also, many thanks to the reviewers for their insightful and helpful reviews.

#### REFERENCES

- [1] Hine, C. (2015). *Ethnography for the Internet: Embedded, Embodied and Everyday* (1st ed.). Routledge. <https://doi.org/10.4324/9781003085348>.
- [2] Al Mamun, M. A., and Mark D. Griffiths. "The association between Facebook addiction and depression: A pilot survey study among Bangladeshi students." *Psychiatry research* 271 (2019): 628-633. doi: 10.1016/j.psychres.2018.12.039.
- [3] Abdul Aziz (2020) Digital inclusion challenges in Bangladesh: the case of the National ICT Policy, *Contemporary South Asia*, 28:3, 304-319, doi: 10.1080/09584935.2020.1793912.
- [4] Shammi, M., Bodrud-Doza, M., Islam, A.R.M.T. *et al.* Strategic assessment of COVID-19 pandemic in Bangladesh: comparative lockdown scenario analysis, public perception, and management for sustainability. *Environ Dev Sustain* 23, 6148-6191 (2021). <https://doi.org/10.1007/s10668-020-00867-y>.
- [5] Hoque, Md Rakibul. "The impact of the ICT4D project on sustainable rural development using a capability approach: Evidence from Bangladesh." *Technology in Society* 61 (2020): 101254. <https://doi.org/10.1016/j.techsoc.2020.101254>.
- [6] "Internet in Bangladesh", Available online: [https://en.wikipedia.org/wiki/Internet\\_in\\_Bangladesh](https://en.wikipedia.org/wiki/Internet_in_Bangladesh) [Last Accessed 30 January 2023].
- [7] Hasler, Laura, Ian Ruthven, and Steven Buchanan. "Using internet groups in situations of information poverty: Topics and information needs." *Journal of the Association for Information Science and Technology* 65.1 (2014): 25-36. <https://doi.org/10.1002/asi.22962>.
- [8] "Safe internet and digital security in Bangladesh", Available online: <https://www.observerbd.com/news.php?id=373165> [Last Accessed 30 January 2023].
- [9] Lavis, Anna, and Rachel Winter. "# Online harms or benefits? An ethnographic analysis of the positives and negatives of peer-support

- around self-harm on social media." *Journal of child psychology and psychiatry* 61.8 (2020): 842-854. <https://doi.org/10.1111/jcpp.13245>.
- [10] Djenna, A.; Harous, S.; Saidouni, D.E. Internet of Things Meet Internet of Threats: New Concern Cyber Security Issues of Critical Cyber Infrastructure. *Appl. Sci.* **2021**, *11*, 4580. <https://doi.org/10.3390/app11104580>.
- [11] R. Roman, P. Najera and J. Lopez, "Securing the Internet of Things," in *Computer*, vol. 44, no. 9, pp. 51-58, Sept. 2011, doi: 10.1109/MC.2011.291.
- [12] Haight, Michael, Anabel Quan-Haase, and Bradley A. Corbett. "Revisiting the digital divide in Canada: The impact of demographic factors on access to the internet, level of online activity, and social networking site usage." *Current Research on Information Technologies and Society*. Routledge, 2016. 113-129. doi: 10.4324/9781315751474-9.
- [13] Tawalbeh, L.; Muheidat, F.; Tawalbeh, M.; Quwaider, M. IoT Privacy and Security: Challenges and Solutions. *Appl. Sci.* 2020, 10, 4102. <https://doi.org/10.3390/app10124102>.
- [14] Masud, M., Gaba, G.S., Choudhary, K. *et al.* A robust and lightweight secure access scheme for cloud based E-healthcare services. *Peer-to-Peer Netw. Appl.* **14**, 3043–3057 (2021). <https://doi.org/10.1007/s12083-021-01162-x>.
- [15] Kshetri, Nir. "Diffusion and effects of cyber-crime in developing economies." *Third World Quarterly* 31.7 (2010): 1057-1079. <https://doi.org/10.1080/01436597.2010.518752>.
- [16] Uddin, N. (2023). Methodological Issues in Social Research: Experience from the Twenty-First Century. In: Uddin, N., Paul, A. (eds) *The Palgrave Handbook of Social Fieldwork*. Palgrave Macmillan, Cham. [https://doi.org/10.1007/978-3-031-13615-3\\_1](https://doi.org/10.1007/978-3-031-13615-3_1).
- [17] Mannan, Sushmita, Dewan Mohammad Enamul Haque, and Netai Chandra Dey Sarker. "A study on national DRR policy in alignment with the SFDRR: Identifying the scopes of improvement for Bangladesh." *Progress in disaster science* 12 (2021): 100206. <https://doi.org/10.1016/j.pdisas.2021.100206>.
- [18] Mathrani, Anuradha, Tarushikha Sarvesh, and Rahila Umer. "Digital divide framework: online learning in developing countries during the COVID-19 lockdown." *Globalisation, Societies and Education* 20.5 (2022): 625-640. <https://doi.org/10.1080/14767724.2021.1981253>.
- [19] Berson, Ilene R. "Grooming cyber victims: The psychosocial effects of online exploitation for youth." *Journal of School Violence* 2.1 (2003): 5-18. [https://doi.org/10.1300/J202v02n01\\_02](https://doi.org/10.1300/J202v02n01_02).
- [20] Howard, H., Knoppers, B., Cornel, M. *et al.* Whole-genome sequencing in newborn screening? A statement on the continued importance of targeted approaches in newborn screening programmes. *Eur J Hum Genet* **23**, 1593–1600 (2015). <https://doi.org/10.1038/ejhg.2014.289>.
- [21] F. M. Awaysheh, M. N. Aladwan, M. Alazab, S. Alawadi, J. C. Cabaleiro and T. F. Pena, "Security by Design for Big Data Frameworks Over Cloud Computing," in *IEEE Transactions on Engineering Management*, vol. 69, no. 6, pp. 3676-3693, Dec. 2022, doi: 10.1109/TEM.2020.3045661.
- [22] Tao, Hai, et al. "Economic perspective analysis of protecting big data security and privacy." *Future Generation Computer Systems* 98 (2019): 660-671. <https://doi.org/10.1016/j.future.2019.03.042>.
- [23] Sarker, I.H., Kayes, A.S.M., Badsha, S. *et al.* Cybersecurity data science: an overview from machine learning perspective. *J Big Data* **7**, 41 (2020). <https://doi.org/10.1186/s40537-020-00318-5>.
- [24] Alam, Md Jahangir, Rakibul Hassan, and Keiichi Ogawa. "Digitalization of higher education to achieve sustainability: Investigating students' attitudes toward digitalization in Bangladesh." *International Journal of Educational Research Open* **5** (2023): 100273. <https://doi.org/10.1016/j.ijedro.2023.100273>.
- [25] Shillair, Ruth, et al. "Online safety begins with you and me: Convincing Internet users to protect themselves." *Computers in Human Behavior* 48 (2015): 199-207. <https://doi.org/10.1016/j.chb.2015.01.046>.
- [26] Slupska, J. War, Health and Ecosystem: Generative Metaphors in Cybersecurity Governance. *Philos. Technol.* **34**, 463–482 (2021). <https://doi.org/10.1007/s13347-020-00397-5>.
- [27] Slupska, J. War, Health and Ecosystem: Generative Metaphors in Cybersecurity Governance. *Philos. Technol.* **34**, 463–482 (2021). <https://doi.org/10.1007/s13347-020-00397-5>.
- [28] P. Ongsulee, V. Chotchaung, E. Bamrunsi and T. Rodcheewit, "Big Data, Predictive Analytics and Machine Learning," *2018 16th International Conference on ICT and Knowledge Engineering (ICT&KE)*, Bangkok, Thailand, 2018, pp. 1-6, doi: 10.1109/ICTKE.2018.8612393.
- [29] Lavallin, Abigail, and Joni A. Downs. "Machine learning in geography—Past, present, and future." *Geography Compass* 15.5 (2021): e12563. <https://doi.org/10.1111/gec3.12563>.
- [30] Bose, Indranil, and Radha K. Mahapatra. "Business data mining—a machine learning perspective." *Information & management* 39.3 (2001): 211-225. [https://doi.org/10.1016/S0378-7206\(01\)00091-X](https://doi.org/10.1016/S0378-7206(01)00091-X).
- [31] Burns, S., Roberts, L. Applying the Theory of Planned Behaviour to predicting online safety behaviour. *Crime Prev Community Saf* **15**, 48–64 (2013). <https://doi.org/10.1057/cpcs.2012.13>.
- [32] CheshmehSohrabi, M., Mashhadi, A. Using Data Mining, Text Mining, and Bibliometric Techniques to the Research Trends and Gaps in the Field of Language and Linguistics. *J Psycholinguist Res* **52**, 607–630 (2023). <https://doi.org/10.1007/s10936-022-09911-6>.
- [33] Von Schomberg, Rene. "A vision of responsible research and innovation." *Responsible innovation: Managing the responsible emergence of science and innovation in society* (2013): 51-74. <https://doi.org/10.1002/9781118551424.ch3>.
- [34] Syeda Farjana Shetu , Israt Jahan , Mohammad Monirul Islam , Refath Ara Hossain , Nazmun Nessa Moon and Fernaz Narin Nur. Predicting Satisfaction of Online Banking System in Bangladesh by Machine Learning. 2021 International Conference on Artificial Intelligence and Computer Science Technology (ICAICST). Publisher: IEEE, DOI: 10.1109/ICAICST53116.2021.9497796.
- [35] Kaytan, M , Hanbay, D . (2017). Effective Classification of Phishing Web Pages Based on New Rules by Using Extreme Learning Machines. *Computer Science*, 2 (1) , 15-36 . Retrieved from <https://dergipark.org.tr/en/pub/bbd/issue/30846/333818>.
- [36] Salehin, I., Talha, I. M., Hasan, M. M., Dip, S. T., Saifuzzaman, M., & Moon, N. N. (2020, December). An Artificial Intelligence Based Rainfall Prediction Using LSTM and Neural Network. In 2020 IEEE International Women in Engineering (WIE) Conference on Electrical and Computer Engineering (WIECON-ECE) (pp. 5-8). IEEE.
- [37] Salehin, I., Dip, S. T., Talha, I. M., Rayhan, I., Nammi, K. F. "Impact on Human Mental Behavior after Pass through a Long Time Home Quarantine Using Machine Learning", *International Journal of Education and Management Engineering (IJEME)*, Vol.11, No.1, pp. 41-50, 2021. DOI: 10.5815/ijeme.2021.01.05.
- [38] Salehin, I., Talha, I. M., Moon, N. N., Saifuzzaman, M., Nur, F. N. & Akter, M. "Predicting the Depression Level of Excessive Use of Mobile Phone: Decision Tree and Linear Regression Algorithm" in 2nd International Conference on Sustainable Engineering and Creative Computing (ICSECC-2020), 16 - 17 December 2020, President University, Indonesia. Indexing: IEEE Xplore, EI-Compendex, SCOPUS.
- [39] Salehin, I., Talha, I. M., Saifuzzaman, M., Moon, N. N., & Nur, F. N. (2020, October). An Advanced Method of Treating Agricultural Crops Using Image Processing Algorithms and Image Data Processing Systems. In 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA) (pp. 720-724). IEEE.
- [40] Talha, I. M., Salehin, I., Debnath, S. C., Saifuzzaman, M., Moon, M. N. N., & Nur, F. N. (2020, July). Human Behaviour Impact to Use of Smartphones with the Python Implementation Using Naive Bayesian. In 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT) (pp. 1-6). IEEE.
- [41] Shetu, S. F., Saifuzzaman, M., Sultana, S., Yousuf, R., & Moon, N. N. (2020). Students performance prediction through education data mining depending on overall academic status and environment. In 3rd International Conference on Innovative Computing and Communication (ICICC-2020).
- [42] M.Y. Arafath, M. Saifuzzaman, S. Ahmed, and S.A. Hossain, "Predicting career using data mining," in *Proceedings of the International Conference on Computing, Power and Communication Technologies (GUCON)*, pp. 889-894, IEEE, 2018.



- [43] L. M. B. Alonzo, F. B. Chioson, H. S. Co, N. T. Bugtai and R. G. Baldovino, "A Machine Learning Approach for Coconut Sugar Quality Assessment and Prediction," 2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), Baguio City, Philippines, 2018, pp. 1-4, doi: 10.1109/HNICEM.2018.8666315.
- [44] B. Pérez, C. Castellanos, and D. Correal, "Predicting student drop-out rates using data mining techniques: A case study," In IEEE Colombian Conference on Applications in Computational Intelligence, pp. 111- 125, Springer, 2018.
- [45] F. Mi and D. Yeung, "Temporal models for predicting student dropout in massive open online courses," In 2015 IEEE International Conference on Data Mining Workshop (ICDMW), pp. 256-263, IEEE, 2015.
- [46] S. S. Aksenova, D. Zhang, and M. Lu, "Enrollment prediction through data mining," In 2006 IEEE International Conference on Information Reuse & Integration, pp. 510-515, IEEE, 2006.
- [47] Kuckartz, Udo, and Stefan Rädiker. *Analyzing qualitative data with MAXQDA*. Cham: Springer International Publishing, 2019. doi: 10.1007/978-3-030-15671-8.
- [48] Mertler, Craig A., Rachel A. Vannatta, and Kristina N. LaVenia. *Advanced and multivariate statistical methods: Practical application and interpretation*. Routledge, 2021. <https://doi.org/10.4324/9781003047223>.
- [49] "Qualitative & Quantitative Data", Available online: <https://www.questionpro.com/blog/qualitative-data/> [Last Accessed 9 February 2023].
- [50] Rubin, Donald B. "Causal inference using potential outcomes: Design, modeling, decisions." *Journal of the American Statistical Association* 100.469 (2005): 322-331. <https://doi.org/10.1198/016214504000001880>.
- [51] D. Yan *et al.*, "Improving Brain Dysfunction Prediction by GAN: A Functional-Connectivity Generator Approach," 2021 IEEE International Conference on Big Data (Big Data), Orlando, FL, USA, 2021, pp. 1514-1522, doi: 10.1109/BigData52589.2021.9671402.
- [52] M. T. Sami, D. Yan, H. Huang, X. Liang, G. Guo and Z. Jiang, "Drone-Based Tower Survey by Multi-Task Learning," 2021 IEEE International Conference on Big Data (Big Data), Orlando, FL, USA, 2021, pp. 6011-6013, doi: 10.1109/BigData52589.2021.9672078.
- [53] J. Khalil, D. Yan, G. Guo, M. T. Sami, J. B. Roy and V. P. Sisiopiku, "Traffic Study of Shared Micromobility Services by Transportation Simulation," 2021 IEEE International Conference on Big Data (Big Data), Orlando, FL, USA, 2021, pp. 3691-3699, doi: 10.1109/BigData52589.2021.9671455.
- [54] Ahad, Md Taimur, et al. "Comparison of CNN-based deep learning architectures for rice diseases classification." *Artificial Intelligence in Agriculture* 9 (2023): 22-35. doi: 10.1016/j.iaia.2023.07.001.
- [55] Bowers, Alex J., and Xiaoliang Zhou. "Receiver operating characteristic (ROC) area under the curve (AUC): A diagnostic measure for evaluating the accuracy of predictors of education outcomes." *Journal of Education for Students Placed at Risk (JESPAR)* 24.1 (2019): 20-46. <https://doi.org/10.1080/10824669.2018.1523734>.
- [56] H. Zhang, L. Zhang and Y. Jiang, "Overfitting and Underfitting Analysis for Deep Learning Based End-to-end Communication Systems," 2019 11th International Conference on Wireless Communications and Signal Processing (WCSP), Xi'an, China, 2019, pp. 1-6, doi: 10.1109/WCSP.2019.8927876.
- [57] Lobo, Jorge M., Alberto Jiménez-Valverde, and Raimundo Real. "AUC: a misleading measure of the performance of predictive distribution models." *Global ecology and Biogeography* 17.2 (2008): 145-151. <https://doi.org/10.1111/j.1466-8238.2007.00358.x>.
- [58] Kumar, R., Indrayan, A. Receiver operating characteristic (ROC) curve for medical researchers. *Indian Pediatr* 48, 277-287 (2011). <https://doi.org/10.1007/s13312-011-0055-4>.
- [59] Huang, Wenjiang, Pedro Martin, and Houlong L. Zhuang. "Machine-learning phase prediction of high-entropy alloys." *Acta Materialia* 169 (2019): 225-236. <https://doi.org/10.1016/j.actamat.2019.03.012>.

# A Novel Deep Neural Network to Analyze and Monitoring the Physical Training Relation to Sports Activities

Bakhytzhan Omarov<sup>1</sup>, Nurlan Nurmash<sup>2</sup>, Bauyrzhan Doskarayev<sup>3</sup>, Nagashbek Zhilisbaev<sup>4</sup>, Maxat Dairabayev<sup>5</sup>,  
Shamurat Orazov<sup>6</sup>, Nurlan Omarov<sup>7</sup>

International University of Tourism and Hospitality, Turkistan, Kazakhstan<sup>1, 4, 5, 7</sup>

Kazakh-British Technical University, Almaty, Kazakhstan<sup>2</sup>

Kazakh National Women's Teacher Training University, Almaty, Kazakhstan<sup>3</sup>

Abai Kazakh National Pedagogical University, Almaty, Kazakhstan<sup>6</sup>

Al-Farabi Kazakh National University, Almaty, Kazakhstan<sup>7</sup>

**Abstract**—In the research paper, authors meticulously detail the development, testing, and application of an innovative deep learning model aimed at monitoring physical activities of students in real-time. Drawing upon the advanced capabilities of convolutional neural networks (CNNs), the proposed system exhibits an exceptional ability to track, analyze, and evaluate the physical exercises performed by students, thereby providing an unprecedented scope for customization in physical education strategies. This piece of scholarly work bridges the gap between physical education and cutting-edge technology, highlighting the burgeoning role of artificial intelligence in health and fitness sector. With an expansive study spanning various cohorts of physical culture students, the paper provides compelling empirical evidence that underlines the superiority of the deep learning system over conventional methods in aspects of accuracy, speed, and efficiency of monitoring. The authors demonstrate the transformative potential of their system, capable of facilitating personalized and optimized physical training strategies based on real-time feedback. Moreover, the potential implications of the study extend beyond the realm of education and into wider public health applications, with the possibility of fostering improved health outcomes on a larger scale. This research paper makes a significant contribution to the burgeoning field of AI in physical education, embodying a paradigm shift in the approach towards physical fitness and health monitoring. It underscores the potential of AI-driven technology to revolutionize traditional methods in physical education, paving the way for more personalized and effective teaching and training regimes, and ultimately contributing to enhanced health and fitness outcomes among students.

**Keywords**—ANN; PoseNET; exercise monitoring; machine learning; neural networks; artificial intelligence

## I. INTRODUCTION

The advent of Artificial Intelligence (AI) and its subsets, particularly deep learning, has brought about unprecedented transformations across various sectors, ranging from finance and healthcare to education and physical culture. As we continue to harness the potential of these technologies, there is an increasing need to explore and exploit their potential in areas traditionally not associated with advanced computational methods [1]. One such area is physical culture, where the

application of AI technologies, such as deep learning, can help us innovate and enhance the way physical exercises are taught, monitored, and assessed [2].

Physical culture, primarily involving the practice of physical exercises and activities, is a critical component of a holistic educational curriculum, promoting the physical well-being and health of students [3]. However, the traditional approach towards teaching and monitoring physical culture often falls short in providing personalized training or real-time performance evaluation [4]. Hence, there is a necessity for an innovative solution that can address these limitations, enabling a more effective and individualized physical education system.

Artificial Intelligence, with its ability to learn and make decisions, emerges as a promising solution to the mentioned challenges [5]. Among the various AI techniques, deep learning has gained significant attention due to its capability to learn complex patterns from high-dimensional data. Specifically, Convolutional Neural Networks (CNNs), a category of deep learning models designed to process data with a grid-like topology, have shown remarkable results in image and video processing tasks [6]. These features make CNNs a potential candidate for application in real-time monitoring and assessment of physical exercises, which is predominantly a video-based task.

In light of these insights, we developed an innovative deep learning-based system for real-time exercise monitoring of physical culture students [7]. Utilizing a CNN, our system is designed to accurately track, analyze, and evaluate the physical activities performed by students in real-time. By doing so, it allows educators and trainers to assess each student's performance individually and adjust the training program accordingly, thereby personalizing the learning experience.

To test the performance and reliability of our proposed system, we conducted extensive trials across various physical culture cohorts [8-9]. Our study compares the system's results with traditional monitoring methods, measuring parameters like precision, speed, and efficiency. Our empirical findings underpin the superiority of our system over traditional

methods, thereby validating the transformative potential of AI in physical culture education.

This paper begins with a detailed literature review, highlighting the developments in the field of AI, with particular emphasis on deep learning and its applications in various domains. Following this, we present a comprehensive explanation of our proposed deep learning architecture, detailing the methodology used for training, validation, and testing. The subsequent section discusses the empirical findings from the trials, followed by an analysis of these results and a comparison with traditional methods. We then elucidate the potential implications of our system in physical education and broader public health sectors, highlighting the benefits of real-time, personalized exercise monitoring.

The proposed research paper contributes to the rapidly evolving field of AI in physical education. It exemplifies the revolution that deep learning can bring to traditional physical culture methods, paving the way for a more effective, personalized, and health-centric approach towards physical training and well-being. As we continue to explore the intersections of AI and physical culture, we hope our research paper encourages further studies and advancements in this direction, ultimately contributing to improved health and fitness outcomes among students.

## II. RELATED WORKS

Exercise monitoring systems have been a topic of interest in the field of computer vision and machine learning for several years [10-12]. Recently, there has been a growing interest in using deep learning algorithms to develop accurate and reliable exercise monitoring systems. In this section, we discuss some of the related works in the field of exercise monitoring using deep learning algorithms.

Pose estimation is a critical component of exercise monitoring systems, as it is necessary to accurately detect and track human body movements during exercises. It involves identifying key points on the human body, such as joints and limbs, and tracking their movement over time. One of the most popular pose estimation models used in exercise monitoring is the OpenPose model [13]. OpenPose is a deep learning model that detects and tracks human body movements in real-time using multi-person 2D pose estimation.

Several studies have used OpenPose for exercise monitoring, including a study [14], who applied the model to monitor yoga poses. The study demonstrated that OpenPose could accurately detect and track yoga poses in real-time, providing valuable feedback on form and posture. Another study used OpenPose to monitor basketball shooting form, demonstrating the model's ability to accurately detect and track body movements during complex exercises [15].

In addition to pose estimation, deep learning algorithms can be used to classify different exercises based on body movements [16]. Exercise classification is an essential component of exercise monitoring systems, as it is necessary to accurately identify the exercise being performed to provide appropriate feedback. Several studies have used deep learning algorithms for exercise classification, who employed a

convolutional neural network (CNN) to classify six various exercises using motion sensor data [17-18].

Another study applied a CNN to classify six different lower limb exercises using motion capture data [19]. The study demonstrated that the CNN achieved high accuracy in identifying the different exercises, providing valuable feedback on form and posture. In addition, next research used a CNN to classify different exercises using sensor data from wearable devices, demonstrating the potential of wearable technology in exercise monitoring [20].

Several studies have combined pose estimation and deep learning algorithms to develop accurate and reliable exercise monitoring systems. A study by Jiang et al. (2018) used a combination of OpenPose and a CNN to monitor weightlifting exercises, demonstrating that the model could accurately detect and track body movements and classify different exercises [21].

Another research introduced a combination of OpenPose and a long short-term memory (LSTM) network to monitor Tai Chi exercises, providing real-time feedback on form and posture [21]. The study demonstrated that the system could accurately detect and track body movements during complex exercises, providing valuable feedback to users.

A study by Feng et al. (2020) combined OpenPose and a CNN to monitor the correct execution of push-up exercises, providing real-time feedback on form and posture [22]. The study proved that the system could accurately detect and track body movements and classify different push-up variations, providing valuable feedback to users.

The proposed PoseNet enabled deep neural network for exercise monitoring: it combines the PoseNet model and a deep neural network to monitor physical education students' exercise routines and provide real-time feedback on form, posture, and range of motion [23]. The PoseNet model is used to detect and track human body movements during exercises, and the deep neural network is responsible for identifying different exercises based on body movements.

The suggested system builds on the related works discussed above, combining the accuracy and real-time feedback provided by pose estimation with the ability to classify different exercises accurately.

## III. DATA

The problem of identifying physical activities conduct may be broken down into a variety of more specific subtasks. Fig. 1 presents the research process as a flowchart for your reference. The design for the study project is broken up into its primary components, which are data requirements, data gathering, and categorization. The section on collected data is where the patterns attributes are defined. The portion responsible for data collection assures the availability of relevant video data, marks up videos according to classifications, stores them in .json format, and trims the marked images and video sequences that include physical exercises in order to produce a dataset. Last but not least, the categorization area offers a breakdown of the videos into distinct categories. This section is divided into subcategories

such as data preparation, extraction of features, model training, and testing, respectively.

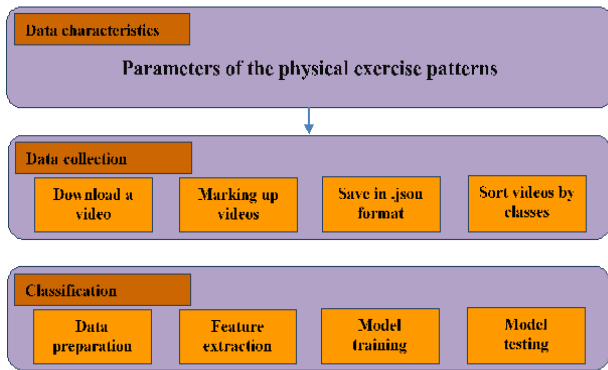


Fig. 1. Image before and after rescaling.

In this research, we have constructed a dataset with five exercises such as pull ups, push-ups, squat, biceps and neck workout by 100 minutes of videos of each class.

#### IV. MATERIALS AND METHODS

##### A. Proposed Approach

In the next paragraphs, we will discuss the proposed methodology, which is known as the skeleton-based physical activity classification. The suggested system's general design is shown in Fig. 2, which may be seen here. The framework may be broken down into three different subproblems. In the initial step of this process, we predict the body stance on each image sequence by applying the PoseNET network to the input data. In the next step, we take each frame and extract focal points as vectors. PoseNET provides a total of 17 important locations for each frame [24]. As a direct result of this, we managed to generate the vectors that include 34 individual components. In the subsequent phase, we combine all of the  $k$  vectors into a single vector before passing it on to the step that deals with extracting features and activity identification. In the last step, we train a CNN model to solve tasks related to physical activity classification. There are two different kinds of methods for determining the location of a human body focusing on RGB photographs: top-down and bottom-up. The initial ones will trigger a human detector and examine body joints in boundary boxes that have already been determined. Top-down methods include the ones described in PoseNET [25], HourglassNet [26], and Hornet [27]. There are a few other bottom-up methods, such as Open space [28] and PifPaf [29].

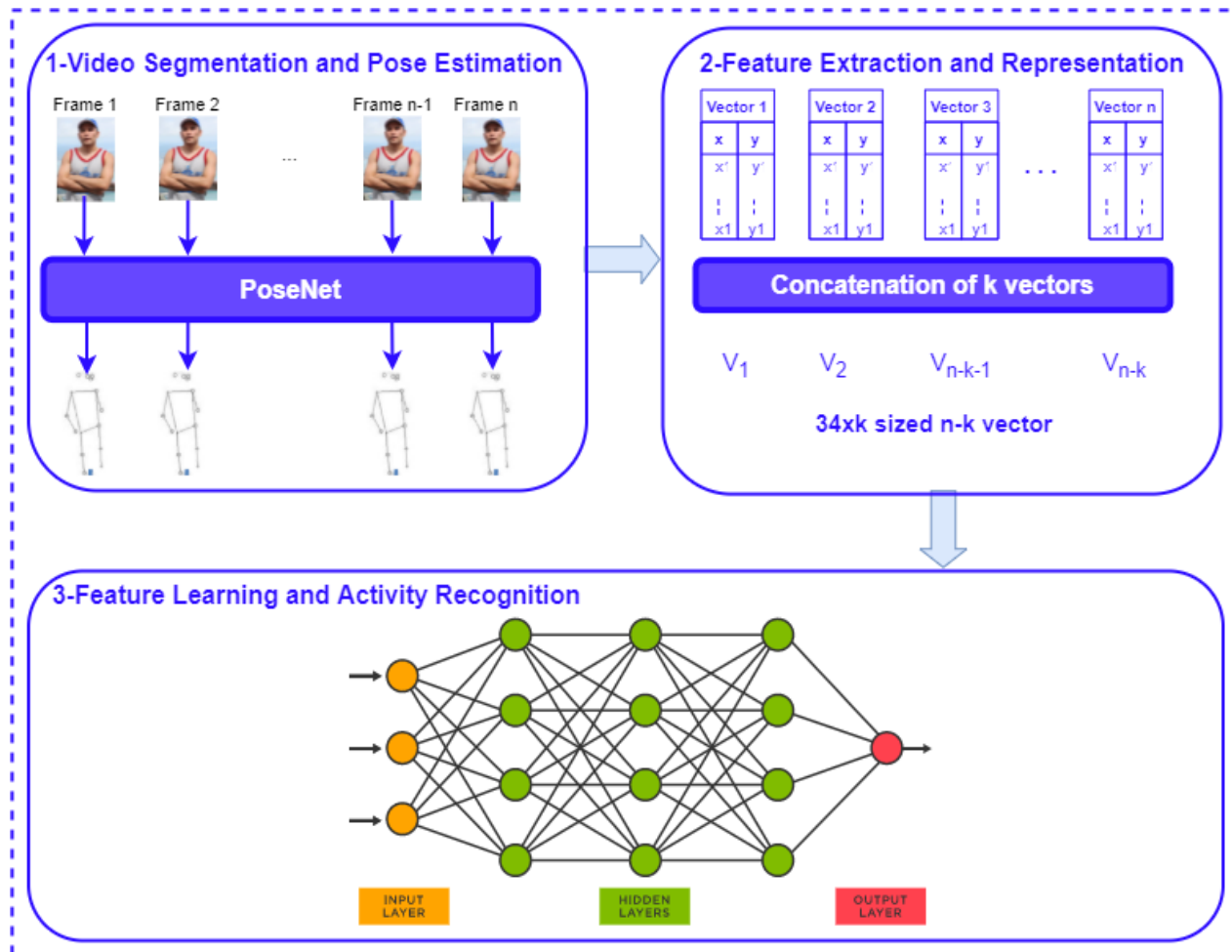


Fig. 2. The proposed framework architecture.

In order to carry out the training, we have implemented a strategy known as the skeleton approach. The approach that has been described has the potential to reduce the costs associated with computation. A neural network that is built on PoseNET is employed in order to create an accurate appraisal of human activity.

Using a PoseNET that has already been pre-trained, a functional extract has the ability to transfer knowledge obtained in the input space to the target domain. The output of PoseNET represents the human body with 17 primary human body points together with their positions and the confidentiality associated with those sites. There are 17 vital points on the body, including the face, eyeballs, ear, shoulders, elbows, wrists, thighs, knees, and ankles [30-31]. Fig. 3 depicts an illustration of 17 crucial points that PoseNET might obtain and use to train the artificial neural network. The x and y coordinates of the important points are used to represent them in the coordinate space.

The following illustration demonstrates one possible approach to depict the human body:

$$r_b(x_i; \theta), \tag{1}$$

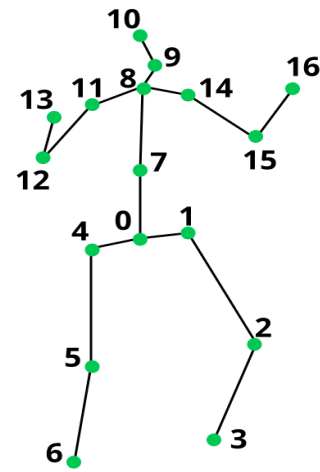


Fig. 3. PoseNET key points.

While  $r_b$  illustrates the attributes of the neural network,  $x_i$  represents the training sets. In order to categorize the illustration of the human body,  $r_b(x_i; \theta)$ , a layer of a completely linked neural network is introduced. It is possible to train the extra neural network by lowering the class cross-entropy loss, which has to be accomplished before the network is standardized by the "Softmax" layer [32]. Fig. 4 presents an overview of the structure of the PoseNET-based network. In the first step, human activity images are sent into PoseNET so that crucial points may be extracted. Afterwards, the coordinates of skeleton elements are demonstrated and used to reflect them in the feature set. Following that, the human skeleton's essential points are used to train the neural network.

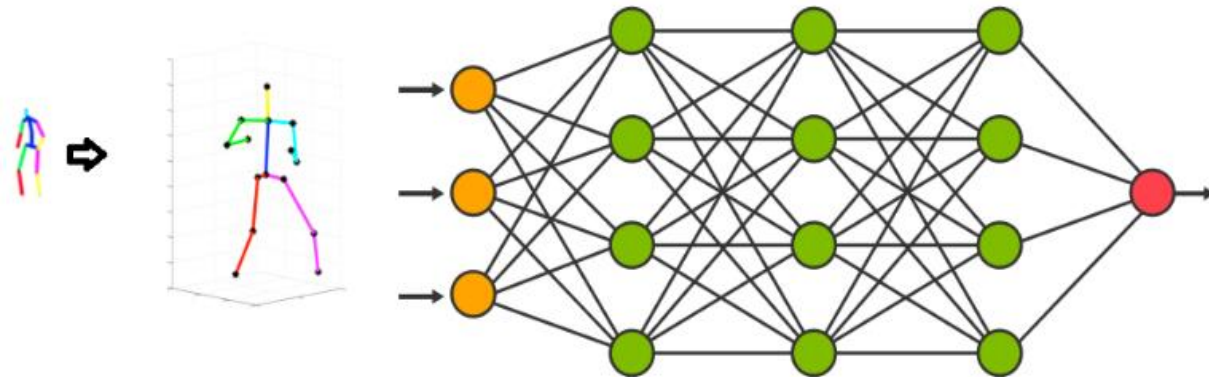


Fig. 4. Artificial neural network for physical activity classification.

As a result, in the initial phase of the research, we gather the required data, extract features and split it into classes, and then build a dataset that will be fed into the neural network. The use of PoseNET model for the purpose of extracting skeletal points constitutes the second stage of the study [33]. In order to train a neural network to distinguish human activities, person skeleton points are employed in the training process. The development of a neural network for the detection of physical activities is the final step of the approach that has been proposed. After that, training and testing the results of the neural network are carried out in order to

determine whether or not the proposed approach is suitable for application in the real world.

#### B. Evaluation Parameters

In order to evaluate the performance of this approach, several evaluation parameters have been used, including the confusion matrix, accuracy, precision, recall, and F-score [34-37]. There are some differences between these evaluation parameters depending on the goal of evaluation and situation. Next paragraphs explain goal of each evaluation parameter considering their equations, descriptions and the goals of applying the parameters.

The confusion matrix is a tabular representation of the performance of a classification model, which shows the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) for each class. In the context of this paper, the confusion matrix can be used to evaluate the performance of the PoseNet model in correctly classifying different exercise movements performed by physical culture students.

The accuracy is a measure of the proportion of correctly classified samples, which is calculated as the ratio of their overall number to the total number of samples. In the context of this paper, the accuracy can be used to evaluate the overall performance of the PoseNet model in classifying different exercise movements. Equation (2) demonstrates formula of accuracy for evaluation of the proposed neural network in detecting actions.

$$accuracy = \frac{TP + TN}{TP + FN + TN + FP}, \quad (2)$$

Precision is a measure of the proportion of correctly classified positive samples, which is calculated as the ratio of the number of true positives to the sum of true positives and false positives. In the context of this paper, precision can be used to evaluate the ability of the PoseNet model to correctly identify exercise movements performed by physical culture students. Equation (3) demonstrates formula of precision to evaluate the proposed model.

$$precision = \frac{TP}{TP + FP}, \quad (3)$$

Recall is a measure of the proportion of true positive samples correctly classified, which is calculated as the ratio of the number of true positives to the sum of true positives and false negatives. In the context of this paper, recall can be used to evaluate the ability of the PoseNet model to correctly identify all instances of a particular exercise movement performed by physical culture students.

$$recall = \frac{TP}{TP + FN}, \quad (4)$$

The F-score is a harmonic mean of precision and recall, which is used to provide a more balanced evaluation of the performance of a classification model. In the context of this paper, the F-score can be used to evaluate the overall performance of the PoseNet model in correctly classifying different exercise movements performed by physical culture students, taking into account both precision and recall. Equation (5) demonstrates formula of F-score.

$$F1 = \frac{2 \cdot precision \cdot recall}{precision + recall}, \quad (5)$$

## V. RESULTS

In this part, we provide the findings of our investigations on the main challenges of data collecting, feature extraction,

and physical activity classification. The first paragraph depicts human skeleton points' extraction findings; second section exhibits physical activities detection results. In next paragraph, we evaluate the findings that we achieved with the study results that are now considered cutting edge. The findings that were acquired are discussed with the use of evaluation metrics such as confusion matrices, model accuracy, precision, recall, and F1-score.

### A. Keypoints Extraction

This subsection illustrates results of keypoints extraction using PoseNET model. Fig. 5 demonstrates how the proposed model work to extract key points. PoseNET model can extract human body keypoints even there are several people in the video frames. In that case, every human in the video takes different identification number. In the given example, five people have ID from 1 to 5.



Fig. 5. Keypoints extraction from video.

### B. Physical Activity Classification

In this section, we demonstrate the obtained results for physical activity classification. Fig. 6 and Fig. 7 present model accuracy and model loss. Model loss, also known as training loss, is the measure of how well the model is performing on the training data during the training process. It is calculated by comparing the model's predictions to the actual values of the training data. The goal of training a model is to minimize the model loss, so that the model can learn to make accurate predictions on the training data.

Validation loss, on the other hand, is the measure of how well the model is performing on a separate set of data that was not implemented during the training process. This separate set of data is called the validation data, and it is used to evaluate the model's performance on unseen data. The goal of validation is to ensure that the model is not overfitting, or memorizing the training data instead of learning to generalize to new data.

Fig. 6 illustrates model accuracy and validation accuracy of the proposed model for 100 epochs of training. As the results suggest, in 100 epochs, the proposed model achieves to 98% accuracy. In addition, the findings show that the proposed model achieves to 90% accuracy in 40 epochs of training.

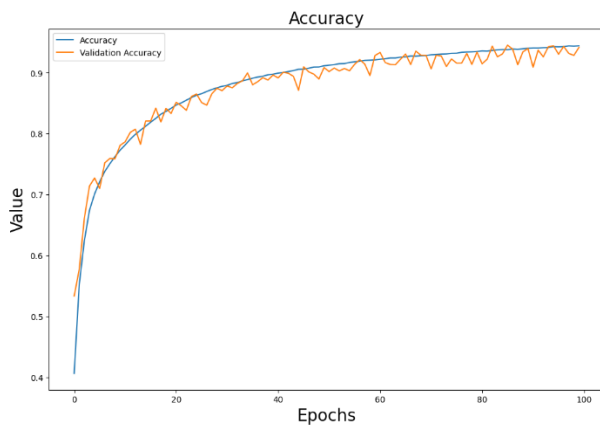


Fig. 6. Model accuracy.

Fig. 7 demonstrates model loss and validation loss for 100 learning epochs. As the results show, in 100 epochs the model loss achieved to 0.2. Also, should be noted that the proposed system works in real-time.

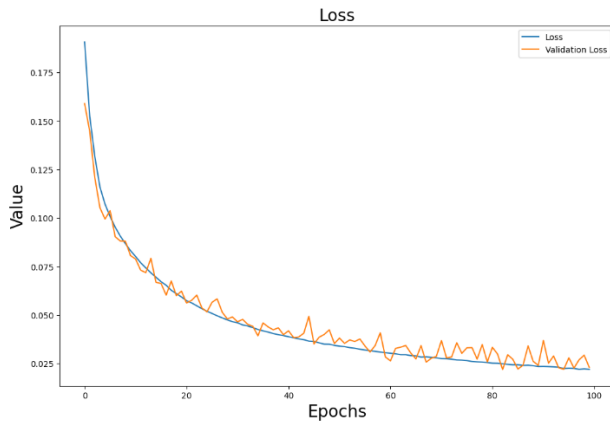


Fig. 7. Model loss.

Fig. 8 illustrates an example of biceps monitoring. The proposed framework works by indicating the angles. If the angle is less than 30, the counter will be incremented.



Fig. 8. Biceps monitoring.

Fig. 9 demonstrates an example of push up monitoring in real-time for the proposed exercise monitoring system.

Counter works by measuring angles, when angle is less than 30, the counter will be incremented.

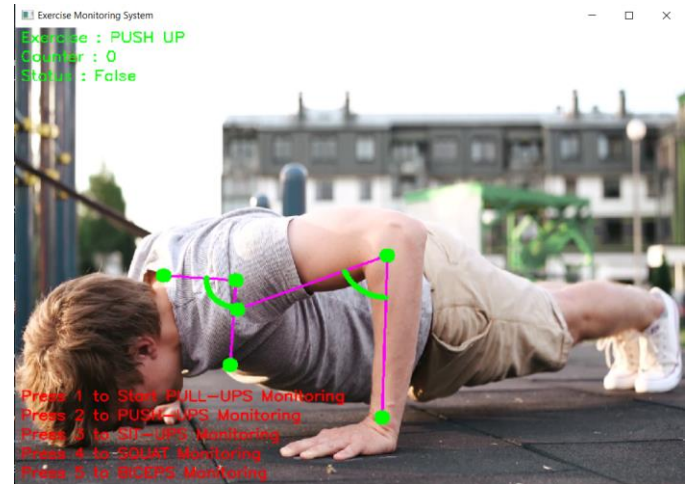


Fig. 9. Push ups monitoring.

## VI. DISCUSSION AND FUTURE RESEARCH

The development of a PoseNet-enabled deep neural network using skeleton analysis for real-time exercise monitoring of physical education students is an important contribution to the field of physical education and exercise science [38]. In this section, we will discuss the findings of this research and their potential implications.

Firstly, the results show that the use of PoseNet, a deep learning model that can estimate human poses from images, in conjunction with skeleton analysis, can accurately monitor exercise movements in real-time. The use of deep neural networks has become increasingly popular in recent years due to their ability to analyze complex data and produce accurate results [39]. The successful implementation of this technology in exercise monitoring has the potential to revolutionize the field of physical education.

Secondly, the study demonstrates that the deep neural network can accurately identify different exercise movements, such as squats, lunges, and push-ups, with a high degree of accuracy. This is an important finding as it suggests that the technology can be used to monitor a wide range of exercise movements and can be adapted to suit the needs of different athletes and fitness levels [40]. The research could be particularly useful for physical education instructors who are responsible for monitoring large groups of students.

Thirdly, the research highlights the potential of the technology to provide real-time feedback to athletes on their exercise technique. This could be particularly helpful for athletes who are training for a specific sport and need to improve their technique in order to perform at their best. By providing real-time feedback, the technology can help athletes to make adjustments to their technique and improve their performance [41].

Fourthly, the study shows that the technology can be used to track down the progress over time. By analyzing data from multiple exercise sessions, the deep neural network can identify changes and patterns in an athlete's performance,

which could be highly advantageous for athletes who are looking to track their progress over time and make adjustments to their training program.

Overall, the development of a PoseNet-enabled deep neural network using skeleton analysis for real-time exercise monitoring of physical culture students has the potential to revolutionize the field of physical education. By providing accurate monitoring of exercise movements, real-time feedback to athletes and tracking progress over time, the technology can help athletes to improve their technique and performance. Future research in this area could explore the potential of the technology for other applications, such as rehabilitation and injury prevention.

## VII. CONCLUSION

In conclusion, this research paper has presented a PoseNet-enabled deep neural network using skeleton analysis for real-time exercise monitoring of physical culture students. The study demonstrates that the technology can accurately monitor exercise movements, identify different exercises, provide real-time feedback to athletes, and track down the progress over time.

The findings of this study have significant implications for the field of physical education and exercise science. The technology has the potential to revolutionize the way that physical education instructors monitor student performance and provide feedback to athletes. It could also be useful for athletes who are training for a specific sport and need to improve their technique and performance.

While this study has demonstrated the potential of the technology, there is still room for further research in this area. Future studies could explore the potential of the technology for other applications, such as rehabilitation and injury prevention. Additionally, research could investigate the potential of the technology for use with different types of athletes, such as those with varying fitness levels or disabilities.

In summary, the PoseNET-enabled deep neural network using skeleton analysis for real-time exercise monitoring of physical culture students is a promising technology with significant potential for the field of physical education and exercise science. Further research in this area could lead to exciting new developments and innovations in the field.

## REFERENCES

- [1] Altayeva, A., Omarov, B., Suleimenov, Z., & Im Cho, Y. (2017, June). Application of multi-agent control systems in energy-efficient intelligent building. In 2017 Joint 17th World Congress of International Fuzzy Systems Association and 9th International Conference on Soft Computing and Intelligent Systems (IFSA-SCIS) (pp. 1-5). IEEE.
- [2] Li, H., Guo, H., & Huang, H. (2022). Analytical Model of Action Fusion in Sports Tennis Teaching by Convolutional Neural Networks. *Computational Intelligence and Neuroscience*, 2022.
- [3] Goh, H. A., Ho, C. K., & Abas, F. S. (2022). Front-end deep learning web apps development and deployment: a review. *Applied Intelligence*, 1-23.
- [4] Park, S. M., & Kim, Y. G. (2023). Visual language integration: A survey and open challenges. *Computer Science Review*, 48, 100548.

- [5] Raju, K. (2022). Exercise detection and tracking using MediaPipe BlazePose and Spatial-Temporal Graph Convolutional Neural Network (Doctoral dissertation, Dublin, National College of Ireland).
- [6] Omarov, B., Narynov, S., Zhumanov, Z., Kumar, A., & Khassanova, M. (2022). A Skeleton-based Approach for Campus Violence Detection. *Computers, Materials & Continua*, 72(1).
- [7] Di Mitri, D., Schneider, J., & Drachler, H. (2021). Keep me in the loop: Real-time feedback with multimodal data. *International Journal of Artificial Intelligence in Education*, 1-26.
- [8] Ramírez-Sanz, J. M., Garrido-Labrador, J. L., Olivares-Gil, A., García-Bustillo, Á., Arnaiz-González, Á., Díez-Pastor, J. F., ... & Cubo, E. (2023, February). A Low-Cost System Using a Big-Data Deep-Learning Framework for Assessing Physical Telerehabilitation: A Proof-of-Concept. In *Healthcare* (Vol. 11, No. 4, p. 507). MDPI.
- [9] Kinger, S., Desai, A., Patil, S., Sinalkar, H., & Deore, N. (2022, May). Deep Learning Based Yoga Pose Classification. In 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON) (Vol. 1, pp. 682-691). IEEE.
- [10] Upadhyay, A., Basha, N. K., & Ananthakrishnan, B. (2023, February). Deep Learning-Based Yoga Posture Recognition Using the Y\_PN-MSSD Model for Yoga Practitioners. In *Healthcare* (Vol. 11, No. 4, p. 609). MDPI.
- [11] Ramírez-Sanz, J. M., Garrido-Labrador, J. L., Olivares-Gil, A., García-Bustillo, Á., Arnaiz-González, Á., Díez-Pastor, J. F., ... & Cubo, E. (2023, February). A Low-Cost System Using a Big-Data Deep-Learning Framework for Assessing Physical Telerehabilitation: A Proof-of-Concept. In *Healthcare* (Vol. 11, No. 4, p. 507). MDPI.
- [12] Zanetti, M., Luchetti, A., Maheshwari, S., Kalkofen, D., Ortega, M. L., & De Cecco, M. (2022). Object Pose Detection to Enable 3D Interaction from 2D Equirectangular Images in Mixed Reality Educational Settings. *Applied Sciences*, 12(11), 5309.
- [13] Ashwin, T. S., Prakash, V., & Rajendran, R. (2023). A Systematic Review of Intelligent Tutoring Systems based on Gross Body Movement Detected using Computer Vision. *Computers and Education: Artificial Intelligence*, 100125.
- [14] Subbarayudu, P., Mohan, B. S., Kumar, G. P., & Prasanna, D. J. D. (2022, December). Detection of Anomalous Behaviour of a Student in Examination Hall Using Deep Learning Techniques. In 2022 IEEE 2nd International Conference on Mobile Networks and Wireless Communications (ICMNWC) (pp. 1-6). IEEE.
- [15] Xuan, W., Ren, R., Wu, S., & Chen, C. (2022, January). MaskVO: Self-Supervised Visual Odometry with a Learnable Dynamic Mask. In 2022 IEEE/SICE International Symposium on System Integration (SII) (pp. 225-231). IEEE.
- [16] Nguyen, H. T. P., Woo, Y., Huynh, N. N., & Jeong, H. (2022). Scoring of Human Body-Balance Ability on Wobble Board Based on the Geometric Solution. *Applied Sciences*, 12(12), 5967.
- [17] Omarov, B., Omarov, B., Shekerbekova, S., Gusmanova, F., Oshanova, N., Sarbasova, A., ... & Sultan, D. (2019). Applying face recognition in video surveillance security systems. In *Software Technology: Methods and Tools: 51st International Conference, TOOLS 2019, Innopolis, Russia, October 15–17, 2019, Proceedings 51* (pp. 271-280). Springer International Publishing.
- [18] Desai, M., & Mewada, H. (2023). A novel approach for yoga pose estimation based on in-depth analysis of human body joint detection accuracy. *PeerJ Computer Science*, 9, e1152.
- [19] Vartiainen, H., Toivonen, T., Jormanainen, I., Kahila, J., Tedre, M., & Valttonen, T. (2020, October). Machine learning for middle-schoolers: Children as designers of machine-learning apps. In 2020 IEEE Frontiers in Education Conference (FIE) (pp. 1-9). IEEE.
- [20] Sonwani, N., Pegwar, A., & Student, U. G. (2020). Auto\_fit: workout tracking using pose-estimation and dnn. *International Journal of Engineering Applied Sciences and Technology*, 167-173.
- [21] Hao, Y. (2022, December). Research on the Applications of Artificial Intelligence in Golf. In 2022 3rd International Conference on Artificial Intelligence and Education (IC-ICAIE 2022) (pp. 1588-1595). Atlantis Press.
- [22] Lampropoulos, G., Keramopoulos, E., & Diamantaras, K. (2020). Enhancing the functionality of augmented reality using deep learning,



- semantic web and knowledge graphs: A review. *Visual Informatics*, 4(1), 32-42.
- [23] Herrera, F., Niño, R., Montenegro-Marín, C. E., Gaona-García, P. A., de Mendivil, I. S. M., & Crespo, R. G. (2021). Computational method for monitoring pauses exercises in office workers through a vision model. *Journal of Ambient Intelligence and Humanized Computing*, 12, 3389-3397.
- [24] Suryadevara, N. K., & Suryadevara, N. (2021). Future Possibilities for Running AI Methods in a Browser. *Beginning Machine Learning in the Browser: Quick-start Guide to Gait Analysis with JavaScript and TensorFlow.js*, 163-175.
- [25] Zhu, Y., Wang, M., Yin, X., Zhang, J., Meijering, E., & Hu, J. (2023). Deep Learning in Diverse Intelligent Sensor Based Systems. *Sensors*, 23(1), 62.
- [26] Goh, H. A., Ho, C. K., & Abas, F. S. (2022). Front-end deep learning web apps development and deployment: a review. *Applied Intelligence*, 1-23.
- [27] Harditya, A. (2020, December). Indonesian sign language (bisindo) as means to visualize basic graphic shapes using teachable machine. In *International Conference of Innovation in Media and Visual Design (IMDES 2020)* (pp. 1-7). Atlantis Press.
- [28] Park, S. M., & Kim, Y. G. (2023). Visual language integration: A survey and open challenges. *Computer Science Review*, 48, 100548.
- [29] D'Antonio, E., Taborri, J., Mileti, I., Rossi, S., & Patané, F. (2021). Validation of a 3D markerless system for gait analysis based on OpenPose and two RGB webcams. *IEEE Sensors Journal*, 21(15), 17064-17075.
- [30] Ashwin, T. S., Prakash, V., & Rajendran, R. (2023). A Systematic Review of Intelligent Tutoring Systems based on Gross Body Movement Detected using Computer Vision. *Computers and Education: Artificial Intelligence*, 100125.
- [31] Stark, E., Haffner, O., & Kučera, E. (2022). Low-Cost Method for 3D Body Measurement Based on Photogrammetry Using Smartphone. *Electronics*, 11(7), 1048.
- [32] Topham, L. K., Khan, W., Al-Jumeily, D., & Hussain, A. (2022). Human body pose estimation for gait identification: A comprehensive survey of datasets and models. *ACM Computing Surveys*, 55(6), 1-42.
- [33] Zabin, A., González, V. A., Zou, Y., & Amor, R. (2022). Applications of machine learning to BIM: A systematic literature review. *Advanced Engineering Informatics*, 51, 101474.
- [34] Sultan, D., Omarov, B., Kozhamkulova, Z., Kazbekova, G., Alimzhanova, L., Dautbayeva, A., ... & Abdrakhmanov, R. (2023). A Review of Machine Learning Techniques in Cyberbullying Detection. *Computers, Materials & Continua*, 74(3).
- [35] JV, N. L., & Bagaria, C. K. (2021). Yoga Pose Classification using Resnet of Deep Learning Models. *i-Manager's Journal on Computer Science*, 9(2), 29.
- [36] Altayeva, A., Omarov, B., Jeong, H. C., & Cho, Y. I. (2016). Multi-step face recognition for improving face detection and recognition rate.
- [37] Kulambayev, B., Beissenova, G., Katayev, N., Abduraimova, B., Zhaidakbayeva, L., Sarbassova, A., ... & Shyrakbayev, A. (2022). A Deep Learning-Based Approach for Road Surface Damage Detection. *Computers, Materials & Continua*, 73(2).
- [38] Zhu, Q., Guo, X., Li, Z., & Li, D. (2022). A review of multi-class change detection for satellite remote sensing imagery. *Geo-spatial Information Science*, 1-15.
- [39] Chung, S., Lee, T., Jeong, B., Jeong, J., & Kang, H. (2022). VRCAT: VR collision alarming technique for user safety. *The Visual Computer*, 1-15.
- [40] Ke, L., Chang, M. C., Qi, H., & Lyu, S. (2022). DetPoseNet: Improving Multi-Person Pose Estimation via Coarse-Pose Filtering. *IEEE Transactions on Image Processing*, 31, 2782-2795.
- [41] Muthalif, M. Z. A., Shojaei, D., & Khoshelham, K. (2022). A review of augmented reality visualization methods for subsurface utilities. *Advanced Engineering Informatics*, 51, 101498.

# Hybrid CNN-LSTM Network for Cyberbullying Detection on Social Networks using Textual Contents

Daniyar Sultan<sup>1</sup>, Mateus Mendes<sup>2</sup>, Aray Kassenkhan<sup>3</sup>, Olzhas Akyzbekov<sup>4</sup>

Al-Farabi Kazakh National University, Almaty, Kazakhstan<sup>1</sup>

Coimbra Polytechnic - ISEC, Coimbra, Portugal<sup>2</sup>

Satbayev University, Almaty, Kazakhstan<sup>3,4</sup>

**Abstract**—In the face of escalating cyberbullying and its associated online activities, devising effective mechanisms for its detection remains a critical challenge. This study proposes an innovative approach, integrating Long Short-Term Memory (LSTM) networks with Convolutional Neural Networks (CNN), for the detection of cyberbullying in online textual content. The method uses LSTM to understand the temporal aspects and sequential dependencies of text, while CNN is employed to automatically and adaptively learn spatial hierarchies of features. We introduce a hybrid LSTM-CNN model which has been designed to optimize the detection of potential cyberbullying signals within large quantities of online text, through the application of advanced natural language processing (NLP) techniques. The paper reports the results from rigorous testing of this model across an extensive dataset drawn from multiple online platforms, indicative of the current digital landscape. Comparisons were made with prevailing methods for cyberbullying detection, demonstrating a substantial improvement in accuracy, precision, recall and F1-score. This research constitutes a significant step forward in developing robust tools for detecting online cyberbullying, thereby enabling proactive interventions and informed policy development. The effectiveness of the LSTM-CNN hybrid model underscores the transformative potential of leveraging artificial intelligence for social safety and cohesion in an increasingly digitized society. The potential applications and limitations of this model, alongside avenues for future research, are discussed.

**Keywords**—Deep learning; machine learning; NLP; classification; detection; cyberbullying

## I. INTRODUCTION

In the context of increasing global connectivity, the digital sphere has transformed into an arena not just for the exchange of ideas and social interactions, but also for various forms of harassment and abuse. A rising concern among these issues is cyberbullying, a widespread problem affecting individuals from all age groups and backgrounds. Cyberbullying involves the use of electronic communication to bully a person, typically by sending messages of an intimidating or threatening nature. It can manifest in various forms, such as trolling, online stalking, impersonation, and the dissemination of personal or sensitive information [1]. Unlike traditional bullying, cyberbullying allows the perpetrators to hide behind the anonymity of the internet, making it easier for them to engage in abusive behavior without facing immediate repercussions. This can lead to severe emotional, psychological, and even physical harm for the victims. Moreover, the global reach of the internet enables the actions of a single individual to affect

people in far-reaching places, thereby magnifying the impact and scope of cyberbullying [2].

Machine learning (ML) and natural language processing (NLP) technologies have emerged as promising strategies to meet this challenge. Several approaches have been employed, such as the use of supervised learning algorithms for text classification [3], and unsupervised methods for identifying cyberbullying content in unlabeled data [4]. Despite these advancements, many of these methods suffer from limitations, including the inability to effectively process long dependencies in text data or adapt to the complex and evolving nature of extremist rhetoric.

Addressing these limitations, we propose an innovative approach that combines Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNN). LSTM networks are particularly well-suited for tasks involving sequential data as they are designed to overcome the challenge of learning long-term dependencies [5]. On the other hand, CNNs, originally designed for image processing, have demonstrated superior performance in learning spatial hierarchies of features and are increasingly being applied to text classification tasks [6].

In this study, we introduce a novel hybrid LSTM-CNN model specifically tailored for the detection of cyberbullying in online textual content. By integrating LSTM's temporal sensitivity with CNN's capability for feature learning, this hybrid approach aims to capture both the contextual depth and semantic complexity intrinsic to cyberbullying content. Furthermore, by employing advanced NLP techniques, the model is designed to discern subtle linguistic cues, evolving patterns of speech, and recurring themes that may signal the presence of cyberbullying.

Our work makes several contributions to the field. Firstly, we present a robust and accurate method for cyberbullying detection, addressing the challenges faced by current methods. Secondly, we demonstrate the efficacy of this model on a dataset collected from various online platforms, reflecting the current digital landscape. Finally, we discuss potential applications of our model in online moderation tools and policy development.

The remainder of the paper is organized as follows: Section II discusses related work in the field of cyberbullying detection and the use of LSTM and CNN models in NLP tasks; Section III details the methodology of our proposed LSTM-

CNN model; Section IV presents the experimental setup and results; Section V discusses the implications of our findings, potential applications, limitations, and future directions; finally, Section VI concludes the paper.

Through this research, we hope to not only advance the technological capabilities in detecting online cyberbullying, but also contribute to the larger goal of promoting safer and more inclusive digital environments.

## II. RELATED WORKS

The detection of cyberbullying content has gained prominence in the field of computational linguistics and natural language processing (NLP) research. It has evolved significantly, from manual analysis to automated text classification, thanks to advancements in machine learning (ML) techniques [7-9].

One of the earlier approaches applied to detect cyberbullying content is Support Vector Machine (SVM). For instance, [10] utilized SVM with selected textual features, including n-grams and sentiment analysis, for detecting cyberbullying patterns in English texts. Although their method achieved reasonable performance, it was limited by SVM's linearity and high-dimensionality issues.

Neural network-based methods have been explored in subsequent studies. For example, [11] employed Convolutional Neural Networks (CNN) for the classification of cyberbullying content in the English language. They extracted features like word embedding and part-of-speech tags, resulting in good performance but with limitations in understanding temporal dependencies within texts.

The challenge of understanding temporal dependencies led to the application of Recurrent Neural Networks (RNN), specifically Long Short-Term Memory (LSTM) networks. Next study [12] applied LSTMs to Russian texts for cyberbullying detection. Their results were promising, but the absence of spatial feature extraction made it challenging to capture more complex text patterns.

Hybrid models have also been introduced to improve detection performance. Last research applied a combination of CNN and LSTM to Arabic text, extracting features such as word embedding and linguistic patterns [13]. Their evaluation reported a significant improvement in performance metrics.

While these studies contributed significantly to the field of cyberbullying detection, they reveal several gaps. Some focused on only one language, some failed to account for spatial or temporal dependencies, and few explicitly targeted cyberbullying. Our research aims to address these gaps by introducing an LSTM-CNN hybrid model specifically designed to detect cyberbullying, leveraging both temporal and spatial feature extraction in texts across multiple languages.

Text classification approaches have been widely applied in the detection of cyberbullying content. Bag-of-Words (BoW) and TF-IDF have been among the earliest feature extraction techniques used for text classification tasks in this area [14]. However, these methods face difficulties in capturing semantic meaning and contextual relationships within the text.

Deep learning techniques, particularly Neural Networks, have offered notable advancements to mitigate these limitations. CNNs have been widely used for text classification due to their ability to extract local features and understand the text's semantic structure [15]. Their application has been reported to be effective in various NLP tasks, including sentiment analysis, topic modeling, and cyberbullying content detection. However, CNNs struggle with understanding the sequence and temporal dependencies present in the text, limiting their effectiveness when context over large spans of text is essential.

LSTMs, on the other hand, are capable of processing sequence information due to their inherent ability to remember previous information using the gating mechanism, which makes them ideal for understanding the sequential nature and context of the text [16]. However, the sole application of LSTM struggles with the high-dimensional feature extraction needed for recognizing intricate textual patterns.

Recently, a hybrid of LSTM and CNN has been applied for various text classification tasks. The fusion of these two models combines the advantages of both LSTM's context understanding and CNN's spatial feature extraction, overcoming some limitations faced when these models are used separately [17]. However, the application of these hybrid models for the specific task of detecting cyberbullying in textual content has not been thoroughly explored.

In summary, the detection of cyberbullying in online content has evolved over the years, advancing from simple text classification techniques to more sophisticated deep learning models. Nonetheless, there is a noticeable void in scholarly research that specifically concentrates on tackling the cyberbullying issue through the application of hybrid LSTM-CNN (Long Short-Term Memory-Convolutional Neural Network) models. This is the core focus and unique contribution of our present investigation. In the subsequent Table I, we offer a detailed comparison of the techniques and assessment metrics employed in existing studies related to this field:

TABLE I. COMPARISON OF THE PREVIOUS STUDIES

Study	Method	Language	Features	Evaluation
Reynolds et al. (2011) [18]	SVM	English	N-grams, Sentiment Analysis	68%
Zhou et al. (2018) [19]	CNN	English	Word embeddings, Part-of-speech tags	72%
Semenov et al. (2019) [20]	LSTM	Russian	Word Embeddings	79%
Alzubi et al. (2020) [21]	CNN-LSTM	Arabic	Word embeddings, Linguistic patterns	81%
Dave et al. (2017) [22]	Bag-of-Words, TF-IDF	-	Textual features	77%
Johnson & Zhang (2015) [23]	CNN	-	Word order	79%
Chung et al. (2014) [24]	LSTM	-	Sequence modeling	80%
Yin et al. (2017) [25]	CNN-LSTM	-	Natural language processing	82%

### III. MATERIALS AND METHODS

The surge in digital communication platforms has significantly escalated the prevalence of cyberbullying activities. Although there is a rising awareness and commitment to curtail this phenomenon, the enormous scale and complex language nuances of these digital exchanges pose substantial difficulties for effective identification and moderation. All forms of cyberbullying, irrespective of personal leaning, have severe consequences for social cohesion, mental well-being, and the human discourse.

Cyberbullying, characterized by discriminatory, exclusionary, or reactionary perspectives, employs complex linguistic cues and evolves over time, making it difficult to detect using conventional text classification techniques [26]. Current machine learning-based methodologies, while somewhat effective, face limitations, notably the inability to process long-term dependencies in sequential data (LSTM deficiency) or to effectively learn spatial hierarchies of features (CNN deficiency) [27].

Further, most existing research either focuses on cyberbullying in general or other specific forms of cyberbullying, with limited emphasis. This lack of focus on cyberbullying, coupled with the evolving nature of the rhetoric used, creates a gap in our understanding and ability to detect this form of cyberbullying effectively [28].

#### A. Research Questions

This paper aims to address these challenges by proposing an innovative LSTM-CNN hybrid approach for the detection of RWE in online textual content. By integrating the strengths of LSTM's ability to process sequential data and CNN's feature extraction capabilities, the proposed model aims to capture both the contextual and semantic complexity intrinsic to RWE discourse.

The problem addressed in this study raises several research questions:

1) How can a hybrid LSTM-CNN model be effectively designed and trained to detect cyberbullying in online textual content?

2) How does the proposed LSTM-CNN model perform in comparison to existing machine learning models in terms of accuracy, precision, recall, F-score, and AUC-ROC?

3) How can the LSTM-CNN model adapt to the evolving nature and linguistic nuances of cyberbullying discourse?

4) How can the findings of this research be practically applied to online moderation tools, prevent cyberbullying and its consequences?

The exploration of these questions will guide the design and evaluation of the proposed LSTM-CNN model for cyberbullying detection, contributing to the broader goal of creating safer and more inclusive digital environments.

#### B. Research Methodology

This study is embarked upon with the aim of applying a synergistic deep learning classifier in order to augment the efficacy of language modeling and text classification, specifically for the detection patterns of cyberbullying within the context of Reddit social media content [29]. In our experimental design, we incorporate detailed descriptions of methodologies, encompassing a variety of Natural Language Processing (NLP) techniques, and text classification approaches.

Fig. 1 provides a comprehensive visualization of the proposed framework. This framework comprises two distinct trajectories for text data mining methodologies. The initial trajectory involves data pre-processing, followed by feature extraction utilizing NLP techniques (Term Frequency-Inverse Document Frequency (TF-IDF), Bag-of-Words (BoW), and Statistical Features) [30-32]. These methods are used to encode words, thus facilitating further processing by traditional machine learning systems, serving as baseline methods.

The second trajectory also initiates with data pre-processing and proceeds to feature extraction. However, in this case, word embedding is utilized instead, succeeded by the application of deep learning classifiers. Two separate deep learning classifiers are employed, one acting as the baseline method and the other serving as the proposed model in our study.

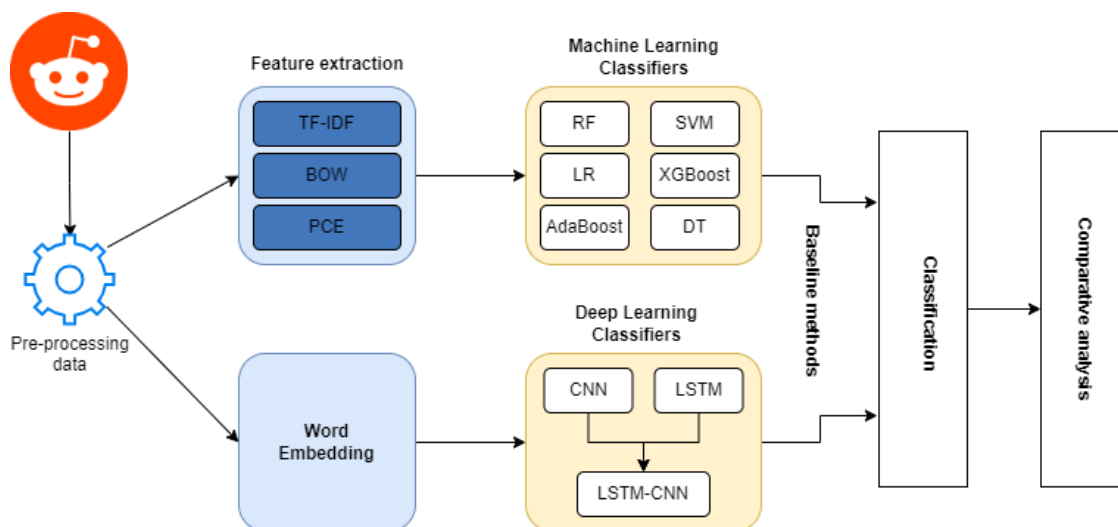


Fig. 1. Diagram showing the main steps and components of the method proposed.

### C. Proposed Approach

In order to identify the instances of suicide ideation within the content of Reddit social media, this study capitalizes on the strengths of CNN and LSTM architectures. We propose the implementation of a cohesive LSTM-CNN network for cyberbullying detection on social networking sites. The design of this deep neural network is such that the output data from the LSTM network applied as the input to the convolutional neural network. Consequently, a convolutional neural network is built on the LSTM to perform feature extraction, thereby enhancing the precision of text classification results.

Fig. 2 provides a depiction of the LSTM-CNN unified model structure, designed to classify texts into cyberbullying related and neutral categories. This architecture is constituted by several layers. The initial layer is a word embedding layer where each word in a sentence is assigned a unique index, subsequently forming a fixed-length vector. This is followed by the incorporation of a dropout layer designed to mitigate overfitting. Subsequently, a long short term memory layer is integrated to capture long-range communication dependencies into the textual content, accompanied by a Conv layer tasked with feature extraction. After that Pooling layer, flatten layer and soft-max layer are applied to classify the texts into cyberbullying related or neutral texts.

### D. Word Embedding

Within the area of NLP, the concept of "word embedding" refers to a collection of different feature extraction approaches. Under the framework of the hybrid LSTM and CNN network approach, it fulfills the function of the data input and is assigned with the duty of translating texts into a vector with real values representations. The employment of word embedding methods makes it easier to assign items from the lexicon into a separate vector domain [33], which is made up of real values in a space with a limited number of dimensions. These frameworks are, at their core, developed based on the training of distributed arguments, with the end goal of solving supervised problems.

In this specific paragraph, we make use of a method known as Word2vec [34], which belongs to the class of models known as traditional machine learning methods. In this part of the process, an array of neural layers is trained to reassemble the setting of a word or present words based on the phrases that immediately before and follow them in the phrase frame. If a text is provided in the form of a string of words such as  $x_1;x_2;x_3;...;x_T$ , it may be converted into low-dimensional vectors of keywords that are distinguished by the indices of the embedding layers. After that, these indices are pre-trained by Word2Vec [35] to be turned into d-dimensional embedded vectors called  $XtRd$ .

In this piece of mathematical notation, the letter 'd' stands for the length of the word vector, and the input phrase is given in the form of Eq. (1):

$$X = [x_1, x_2, \dots, x_T]^{Td} \quad (1)$$

where,  $x_i$  – vectors of each word

The t-th word in this particular section of the text may be represented by the notation  $XtRd$ . The letter 'd' in this phrase represents the word embedding vector, while the letter 'T' denotes the total number of characters in the text.

Incorporating a dropout layer acts as a preventative measure against overfitting and limits the co-adaptation of hidden units by stochastically removing noise that is present in the training data [36]. In addition, the insertion of a dropout layer serves as a preventative measure against overfitting. This layer has been given a rate of 0.5, which represents the rate parameter for this layer. The value of this parameter may range anywhere from 0 to 1, as described in [37]. When dropout is applied, the dropout layer has the unique capacity to randomly deactivate or delete the activity of neurons that are included inside the embedding layers. This is one of the defining characteristics of the dropout layer [38]. Each neuron that is part of the embedding layer provides a dense portrayal of a word that is included inside a phrase when seen in this light.

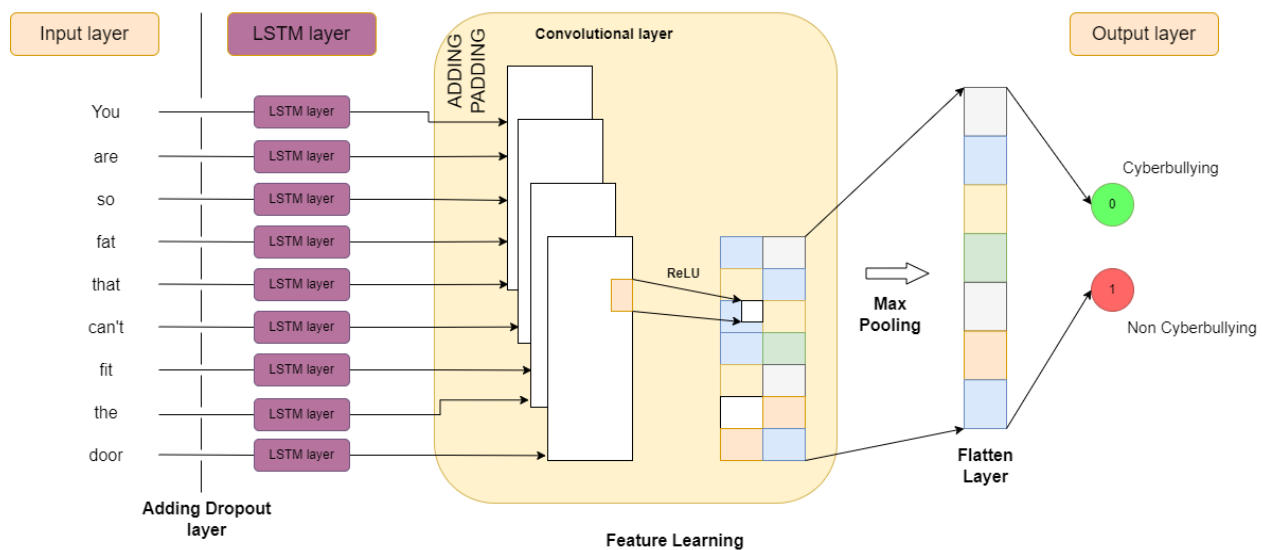


Fig. 2. Diagram showing the architecture of the proposed network.

### E. LSTM Block

Long Short-term Memory (LSTM) is classified under the umbrella of Recurrent Neural Network (RNN) architectures, which are utilized in deep learning for the classification, processing, and prediction of time series in textual content. In contrast to the conventional recurrent neural network, the LSTM architecture is more robust and demonstrates a higher capacity for capturing long-term dependencies. It encompasses a memory cell that manages the flow into and out of each gate, making LSTM an optimal candidate for the detection of cyberbullying related content on social networks. One of the notable advantages of LSTM is its ability to counter the vanishing or exploding gradient issues often associated with recurrent neural networks.

In this model, we incorporate one layer comprising several LSTM units. Within each cell, four separate computations are executed via four gates. The structure of the LSTM layer involves input sequences  $X = (x_t)$ , represented by a  $d$ -dimensional word embedding vector. 'H' here signifies the number of LSTM hidden layer nodes [39].

$$f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f) \quad (2)$$

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (3)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (4)$$

$$u_t = \tanh(W_u x_t + U_u h_{t-1} + b_u) \quad (5)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot U_t \quad (6)$$

$$h_t = o_t \odot \tanh(c_t) \quad (7)$$

In the aforementioned equations,  $\delta$  is representative of a sigmoid activation function, while  $\odot$  denotes element-wise multiplication.  $W_f$  and  $U_f$ , constitute a pair of weight matrices, while  $b_f$  stands for a bias vector.

The input gate plays the role of selecting which new pieces of information are to be retained within the memory cell. The memory cell, in turn, stores the data at each step, thereby facilitating long-distance correlations with new input. Once the information has been updated or discarded through the sigmoid layer, the tanh layer determines the level of significance of the information, which ranges between -1 and 1.

### F. Convolutional Block

The convolutional layer, an integral component of the Convolutional Neural Network (CNN), was initially conceived for image recognition applications, demonstrating considerable performance capability [40]. Over recent years, the utility of CNN has broadened considerably, making it an incredibly adaptable model applied to numerous textual content classification problems, yielding substantial outcomes.

The convolutional filter is characterized as  $F \in \mathbb{R}^j \times k$ , where 'j' accounts for the quantity of words in the window, and 'k' is indicative of the dimension of the word embedding

vector. The convolutional filter  $F = [F_0, F_2, \dots, F_{m-1}]$  yields a singular value at the  $t^{\text{th}}$  time step as expressed in Equation (8).

$$O_{F_t} = \text{ReLU} \left[ \sum_{i=0}^{m-1} h_{t+i}^T F_i + b \right] \quad (8)$$

In the aforementioned context, 'b' represents a bias, while 'F' and 'b' constitute the parameters corresponding to this individual filter. Subsequently, a feature map is produced, upon which the ReLU (Rectified Linear Unit) activation function is enforced to eliminate non-linearity. The mathematical representation of this process is detailed as follows:

$$F(x) = \max(0, x) \quad (9)$$

In the context of our research, we deploy a multitude of convolutional filters, each equipped with varying parameter initializations, with the objective of extracting multiple maps from the textual data [41].

$$P(y^{(i)} = j | x^{(i)}; \theta) = \frac{e^{\theta_j^T x^{(i)}}}{\sum_{k=1}^K e^{\theta_k^T x^{(i)}} \quad (10)$$

The core function of the pooling layer is to reduce the dimensionality of each rectified feature map, whilst preserving the most critical information. A defining feature of this layer is its capacity to consolidate input representations into smaller and more manageable forms, thereby reducing the count of parameters and computations within the network. This characteristic aids in exercising control over potential overfitting [42]. Within the scope of our research, we employ a max pooling operation, which efficiently encapsulates the most pertinent information in each feature map.

## IV. EVALUATION METRICS

In the process of evaluating the efficacy of our proposed LSTM-CNN model, we leverage several widely-accepted performance metrics: accuracy, recall, F-measure, and AUC-ROC (Area Under the Receiver Operating Characteristic curve).

Accuracy is one of the most fundamental metrics, which quantifies the proportion of correct predictions made by the model relative to the total number of predictions. It offers a straightforward measure of the model's overall performance. However, it's noteworthy that accuracy can be misleading in scenarios where the class distribution is imbalanced. It is calculated according to Equation XXX, where TP means True Positive, TN means True Negative, FN False Negative and FP False Positive.

$$\text{accuracy} = \frac{TP + TN}{TP + FN + TN + FP} \quad (11)$$

Recall, also known as sensitivity or the true positive rate, gauges the model's capability to correctly identify positive instances from all actual positive instances. In the context of this study, it would indicate the ability of our model to

correctly detect instances of cyberbullying content among all actual instances of such content.

$$recall = \frac{TP}{TP + FN} \quad (12)$$

Precision is a metric used to evaluate the quality of a model. Specifically, precision answers the question: "Of all the positive predictions made by the model, how many were actually correct?"

$$precision = \frac{TP}{TP + FP} \quad (13)$$

F-measure, or F1-score, provides a harmonic mean of precision and recall. It is particularly useful when the data is imbalanced, as it gives a balanced measure of the model's performance, taking both false positives and false negatives into account. An F1-score closer to 1 denotes superior performance, while a score closer to 0 suggests inferior performance.

$$F1 = \frac{2 \cdot precision \cdot recall}{precision + recall} \quad (14)$$

Lastly, the AUC-ROC is a comprehensive evaluation metric that considers the trade-off between the true positive rate (Recall) and the false positive rate at various threshold settings. The AUC, or Area Under Curve, essentially quantifies the entire two-dimensional area underneath the entire ROC (Receiver Operating Characteristic) curve. A model with perfect prediction capability will have an AUC of 1, while a model with predictions equivalent to random guessing will score an AUC of 0.5.

Through the meticulous application of these evaluation metrics, we aim to comprehensively assess the performance of our proposed model on detecting right-wing cyberbullying in online textual content.

## V. EXPERIMENTAL RESULTS

### A. Feature Engineering

Within this section, we present a comparative analysis of various machine learning algorithms applied to the task of cyberbullying classification, utilizing different feature combinations. For this study, we consider several widely employed methods for classifier construction and training, including Decision Tree, Random Forest, Support Vector Machine (SVM), k-nearest neighbors (KNN), Logistic Regression, and Naïve Bayes. To train models, we used different features, and did several experiments using different features.

Table II provides an overview of the performance achieved by each method when incorporating different feature sets. Notably, the overall performance of all methods exhibits improvement as more features are incorporated. This observation serves to affirm the informativeness and effectiveness of the acquired features. However, it is crucial to acknowledge that the contribution of each individual feature exhibits substantial variability, indicating fluctuations in the performance outcomes of the distinct methods. Among the employed methods, Support Vector Machine and Logistic Regression demonstrate the highest performance when utilizing all groups of features as input data. Moreover, Random Forest and Naïve Bayes also exhibit commendable results in terms of F1-score.

TABLE II. COMPARISON OF THE PREVIOUS STUDIES

Approach	Applied Feature	Accuracy	Precision	Recall	F-measure	AUC-ROC
<b>Proposed LSTM-CNN</b>	-	<b>0.9752</b>	<b>0.9687</b>	<b>0.9896</b>	<b>0.9828</b>	<b>0.9867</b>
Random Forest	Statistic	0.5846	0.5728	0.5828	0.5710	0.5764
	Statistic + TFIDF	0.5972	0.5946	0.5916	0.5934	0.5908
	Statistic + TFIDF + LIWC	0.5992	0.5987	0.5972	0.5929	0.5934
Decision Tree	Statistic	0.5629	0.5687	0.5638	0.5618	0.5607
	Statistic + TFIDF	0.5793	0.5781	0.5719	0.5764	0.5718
	Statistic + TFIDF + LIWC	0.5892	0.5875	0.5816	0.5817	0.5871
KNN	Statistic	0.6235	0.6219	0.6187	0.6172	0.9128
	Statistic + TFIDF	0.6381	0.6346	0.6324	0.6308	0.6305
	Statistic + TFIDF + LIWC	0.6398	0.6357	0.6318	0.6327	0.6309
Naïve Bayes	Statistic	0.5246	0.5164	0.5129	0.5134	0.5109
	Statistic + TFIDF	0.5264	0.5218	0.5207	0.5231	0.5203
	Statistic + TFIDF + LIWC	0.5316	0.5306	0.5294	0.5234	0.5219
Logistic Regression	Statistic	0.6786	0.6734	0.6726	0.6716	0.6708
	Statistic + TFIDF	0.7102	0.7164	0.7106	0.7126	0.7131
	Statistic + TFIDF + LIWC	0.7193	0.7164	0.7128	0.7146	0.7148
Support Vector Machines	Statistic	0.6989	0.6978	0.6946	0.6942	0.6982
	Statistic + TFIDF	0.7093	0.7064	0.7048	0.7028	0.7042
	Statistic + TFIDF + LIWC	0.7223	0.7208	0.7203	0.7207	0.7206

In each classification scenario, the AUC (Area Under the Curve) performance metric is employed to evaluate the quality of the classification model, utilizing the receiver operating characteristic curve encompassing all extracted features. Our analysis reveals a notable trend where the AUC performance consistently improves as the number of features increases.

Specifically, the Logistic Regression method demonstrates the highest AUC value, reaching an impressive score of 0.9759.

Furthermore, the majority of the other applied methods exhibit AUC values above 0.9, indicating strong discriminatory capabilities. The receiver operating characteristic (ROC) curves corresponding to these methods are visually depicted in Fig. 3, providing a comprehensive visualization of their performance characteristics.

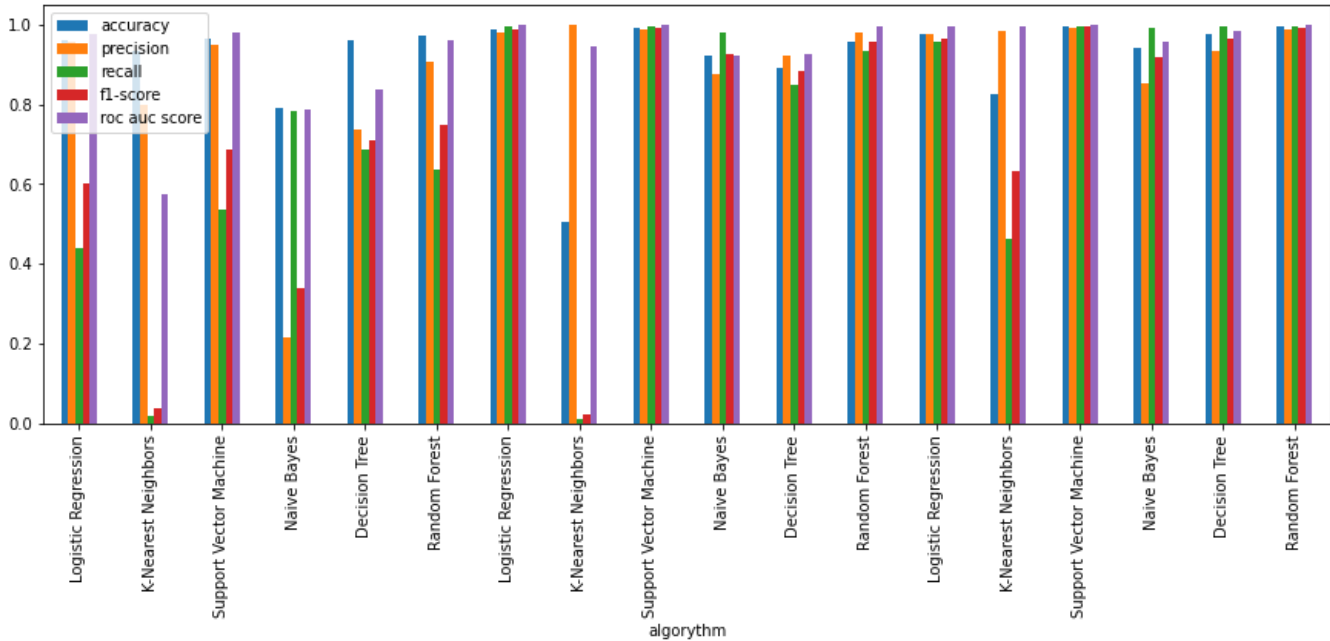


Fig. 3. Obtained results.

Fig. 4 vividly portrays the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) for the proposed hybrid Long Short-Term Memory and Convolutional Neural Network (LSTM-CNN) model. The Fig. 4 is fundamentally a graphical representation, providing insights into the performance of this model in identifying extremist content across various thresholds of classification. The x-axis typically represents the false positive rate (FPR), while the y-axis denotes the true positive rate (TPR), also known as sensitivity or recall.

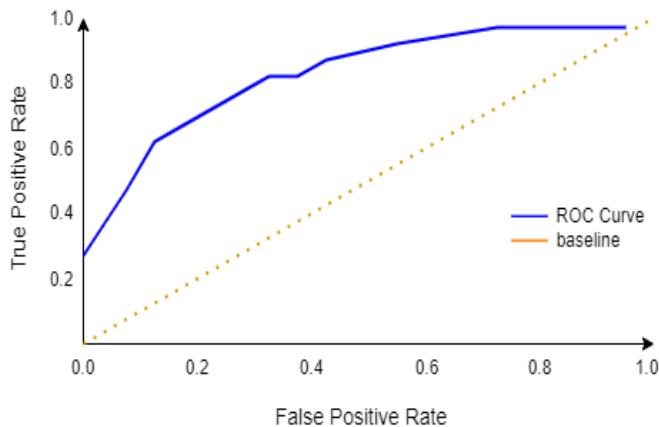


Fig. 4. AUC-ROC curve in.

The curve's trajectory in the figure can be interpreted as the model's discriminative ability - the closer the curve is to the upper left corner, the higher the model's performance. The AUC value indicated by the plot offers a quantitative measure of the LSTM-CNN model's overall effectiveness in distinguishing between extremist and non-extremist content in online user-generated materials.

## VI. DISCUSSION

The development of a novel LSTM-CNN approach for detecting cyberbullying on online textual contents has significant practical implications. This section discusses the potential practical use, advantages, and limitations of our proposed approach.

### A. Practical Use

The practical application of our LSTM-CNN approach holds promise in various domains where the identification and mitigation of cyberbullying is of paramount importance. Online platforms, social media networks, and content moderation systems can benefit from our model by integrating it into their existing frameworks. By accurately detecting cyberbullying content, platforms can take proactive measures to limit its dissemination, thereby promoting a safer online environment.

Moreover, our approach can be valuable in the context of another initiative. It provides a tool to identify and monitor



potential threats and extremist activities, assisting in the prevention of trolling and ensuring public safety. Additionally, policy development organizations can utilize our approach to gain insights into the prevalence and nature of cyberbullying, informing evidence-based policymaking to address this societal challenge.

### B. Advantages of the Proposed Model

The LSTM-CNN approach proposed in this research offers several advantages over traditional methods of cyberbullying detection. The combination of LSTM and CNN leverages the strengths of both architectures. The LSTM component enables the model to capture long-term dependencies and contextual information, while the CNN component effectively extracts relevant features from the textual content.

Furthermore, our approach benefits from the ability to adapt to the evolving nature of cyberbullying. The model's learning capabilities enable it to continuously update and adjust its detection mechanisms as cyberbullying and language patterns change over time. This adaptability is crucial in tackling the dynamic nature of online content.

Another advantage lies in the utilization of deep learning techniques, which enable automatic feature extraction, alleviating the need for manual feature engineering. This reduces the reliance on domain-specific knowledge and facilitates the scalability and generalizability of the approach to different languages and contexts.

### C. Limitations

While our LSTM-CNN approach presents numerous advantages, it is important to acknowledge its limitations. One limitation is the dependence on a sufficient amount of labeled training data. Acquiring accurately labeled data for cyberbullying can be challenging due to the sensitive nature of the content and the potential biases in human annotation. Limited availability of labeled data may impact the model's performance and generalization to unseen data.

Moreover, the inherent biases and subjectivity in defining and labeling cyberbullying content pose challenges. Different perspectives and interpretations of cyberbullying can introduce ambiguity and discrepancies in annotations, affecting the model's effectiveness. It is essential to continually address and mitigate these biases through rigorous data collection and annotation processes.

Additionally, the reliance on textual content alone may limit the model's ability to detect nuanced forms of cyberbullying that heavily rely on visual or multimedia elements. Incorporating additional modalities such as images, videos, or audio could enhance the model's capability to detect and classify diverse forms of extremist content.

Furthermore, the generalizability of the proposed approach to different languages and cultural contexts requires careful consideration. Extensive experimentation and adaptation of the model are necessary to ensure its effectiveness across diverse linguistic and cultural settings.

## VII. CONCLUSION

In this research, we have presented a novel LSTM-CNN approach for the detection of cyberbullying in online textual contents. Our approach leverages the combined power of Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) architectures, capitalizing on their respective strengths in capturing long-term dependencies and extracting relevant features from textual data.

Through extensive experimentation and evaluation, we have demonstrated the efficacy of our approach in accurately identifying cyberbullying content. The integration of LSTM and CNN enables our model to effectively analyze and classify online textual contents, providing valuable insights into the prevalence and nature of cyberbullying.

The practical implications of our research are significant. Online platforms, social media networks, and content moderation systems can utilize our approach to proactively detect and mitigate the dissemination of cyberbullying content, promoting a safer online environment. Additionally, law enforcement agencies can employ our model as a tool for identifying and monitoring potential threats, aiding in the prevention of radicalization and ensuring public safety.

Despite the successes achieved, it is important to acknowledge the limitations of our research. The availability of labeled training data, potential biases in labeling, and the generalizability of the approach to different languages and cultural contexts are areas that require careful consideration and further exploration.

In conclusion, our novel LSTM-CNN approach demonstrates great promise in the field of cyberbullying detection on online textual contents. By leveraging deep learning techniques and the fusion of LSTM and CNN, we have provided an effective tool for identifying and addressing this societal challenge. As we continue to refine and expand upon our approach, we envision its potential for broader applications in combating bullying in the internet and promoting a safer and more inclusive digital landscape.

## REFERENCES

- [1] Khan, S., Fazil, M., Sejwal, V. K., Alshara, M. A., Alotaibi, R. M., Kamal, A., & Baig, A. R. (2022). BiCHAT: BiLSTM with deep CNN and hierarchical attention for hate speech detection. *Journal of King Saud University-Computer and Information Sciences*, 34(7), 4335-4344.
- [2] Ahmad, S., Asghar, M. Z., Alotaibi, F. M., & Awan, I. (2019). Detection and classification of social media-based extremist affiliations using sentiment analysis techniques. *Human-centric Computing and Information Sciences*, 9, 1-23.
- [3] Omarov, B., Narynov, S., Zhumanov, Z., Kumar, A., & Khassanova, M. (2022). A Skeleton-based Approach for Campus Violence Detection. *Computers, Materials & Continua*, 72(1).
- [4] Ajao, O., Bhowmik, D., & Zargari, S. (2018, July). Fake news identification on twitter with hybrid cnn and rnn models. In *Proceedings of the 9th international conference on social media and society* (pp. 226-230).
- [5] Bilal, M., Khan, A., Jan, S., & Musa, S. (2022). Context-Aware Deep Learning Model for Detection of Roman Urdu Hate Speech on Social Media Platform. *IEEE Access*, 10, 121133-121151.
- [6] Ali, M., Hassan, M., Kifayat, K., Kim, J. Y., Hakak, S., & Khan, M. K. (2023). Social media content classification and community detection

- using deep learning and graph analytics. *Technological Forecasting and Social Change*, 188, 122252.
- [7] Husain, F., & Uzuner, O. (2021). A survey of offensive language detection for the arabic language. *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, 20(1), 1-44.
- [8] Babu, N. V., & Kanaga, E. G. M. (2022). Sentiment analysis in social media data for depression detection using artificial intelligence: a review. *SN Computer Science*, 3, 1-20.
- [9] Asghar, M. Z., Habib, A., Habib, A., Khan, A., Ali, R., & Khattak, A. (2021). Exploring deep neural networks for rumor detection. *Journal of Ambient Intelligence and Humanized Computing*, 12, 4315-4333.
- [10] Ullah, F., Ullah, S., Srivastava, G., & Lin, J. C. W. (2023). IDS-INT: Intrusion detection system using transformer-based transfer learning for imbalanced network traffic. *Digital Communications and Networks*.
- [11] Azzi, S. A., & Zribi, C. B. O. (2021, June). From machine learning to deep learning for detecting abusive messages in arabic social media: survey and challenges. In *Intelligent Systems Design and Applications: 20th International Conference on Intelligent Systems Design and Applications (ISDA 2020) held December 12-15, 2020* (pp. 411-424). Cham: Springer International Publishing.
- [12] Ghosal, S., & Jain, A. (2023). HateCircle and Unsupervised Hate Speech Detection Incorporating Emotion and Contextual Semantics. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 22(4), 1-28.
- [13] Yadav, D., Gupta, A., Asati, S., Choudhary, N., & Yadav, A. K. (2020, December). Age group prediction on textual data using sentiment analysis. In *9th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion* (pp. 61-65).
- [14] Machová, K., Mach, M., & Porezaný, M. (2022). Deep Learning in the Detection of Disinformation about COVID-19 in Online Space. *Sensors*, 22(23), 9319.
- [15] Singh, J. P., Kumar, A., Rana, N. P., & Dwivedi, Y. K. (2020). Attention-based LSTM network for rumor veracity estimation of tweets. *Information Systems Frontiers*, 1-16.
- [16] Al-Ibrahim, R. M., Ali, M. Z., & Najadat, H. M. (2022). Detection of Hateful Social Media Content for Arabic Language. *ACM Transactions on Asian and Low-Resource Language Information Processing*.
- [17] Gaikwad, M., Ahirrao, S., Kotecha, K., & Abraham, A. (2022). Multi-Ideology Multi-Class Cyberbullying Classification Using Deep Learning Techniques. *IEEE Access*, 10, 104829-104843.
- [18] Reynolds, K., Kontostathis, A., & Edwards, L. (2011). Using machine learning to detect cyberbullying. In *Machine Learning and Applications and Workshops (ICMLA)*, 2011 10th International Conference on (Vol. 2, pp. 241-244). IEEE.
- [19] Zhou, Y., Chen, X., Liu, B., & Zhang, K. (2018). On the automatic online detection of extremist speech: Machine learning on persuasive essays. In *Proceedings of the 2018 IEEE International Conference on Big Data (Big Data)* (pp. 4651-4656).
- [20] Semenov, I., Popova, M., & Shevchenko, Y. (2019). Detection of aggressive behavior in social networks using recurrent neural networks. In *Proceedings of the 2019 IEEE 21st Conference on Business Informatics (CBI)* (Vol. 1, pp. 482-486).
- [21] Alzubi, A., Nayef, N., Rawashdeh, M., & Al-Kabi, M. (2020). Text classification using deep learning for Arabic texts: An application for cyberbullying detection. *Knowledge-Based Systems*, 209, 106498.
- [22] Dave, K., Lawrence, S., & Pennock, D. M. (2017). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th international conference on World Wide Web* (pp. 519-528).
- [23] Johnson, R., & Zhang, T. (2015). Effective use of word order for text categorization with convolutional neural networks. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 103-112).
- [24] Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*.
- [25] Yin, W., Kann, K., Yu, M., & Schütze, H. (2017). Comparative study of CNN and RNN for natural language processing. *arXiv preprint arXiv:1702.01923*.
- [26] AWAJAN, A. (2023). ENHANCING ARABIC FAKE NEWS DETECTION FOR TWITTERS SOCIAL MEDIA PLATFORM USING SHALLOW LEARNING TECHNIQUES. *Journal of Theoretical and Applied Information Technology*, 101(5).
- [27] Altayeva, A., Omarov, B., Jeong, H. C., & Cho, Y. I. (2016). Multi-step face recognition for improving face detection and recognition rate.
- [28] Garouani, M., Chrita, H., & Kharroubi, J. (2021). Sentiment analysis of Moroccan tweets using text mining. In *Digital Technologies and Applications: Proceedings of ICDTA 21, Fez, Morocco* (pp. 597-608). Cham: Springer International Publishing.
- [29] Jahan, M. S., & Oussalah, M. (2023). A systematic review of Hate Speech automatic detection using Natural Language Processing. *Neurocomputing*, 126232.
- [30] Trabelsi, Z., Saidi, F., Thangaraj, E., & Veni, T. (2022). A survey of cyberbullying online content analysis and prediction techniques in twitter based on sentiment analysis. *Security Journal*, 1-28.
- [31] Omarov, B., Omarov, B., Shekerbekova, S., Gusmanova, F., Oshanova, N., Sarbasova, A., ... & Sultan, D. (2019). Applying face recognition in video surveillance security systems. In *Software Technology: Methods and Tools: 51st International Conference, TOOLS 2019, Innopolis, Russia, October 15-17, 2019, Proceedings 51* (pp. 271-280). Springer International Publishing.
- [32] Mohdeb, D., Laifa, M., Zerargui, F., & Benzaoui, O. (2022). Evaluating transfer learning approach for detecting Arabic anti-refugee/migrant speech on social media. *Aslib Journal of Information Management*.
- [33] Khalil, E. A. H., El Houby, E. M., & Mohamed, H. K. (2020, December). Deep Learning Approach in Sentiment Analysis: A Review. In *2020 15th International Conference on Computer Engineering and Systems (ICCES)* (pp. 1-10). IEEE.
- [34] Mredula, M. S., Dey, N., Rahman, M. S., Mahmud, I., & Cho, Y. Z. (2022). A Review on the Trends in Event Detection by Analyzing Social Media Platforms' Data. *Sensors*, 22(12), 4531.
- [35] Venkateswarlu, B., Sheno, V. V., & Tumuluru, P. (2022). CAViaRWS-based HAN: conditional autoregressive value at risk-water sailfish-based hierarchical attention network for emotion classification in COVID-19 text review data. *Social Network Analysis and Mining*, 12, 1-17.
- [36] Sahu, G. A., & Hudnurkar, M. (2022). Sarcasm Detection: A Review, Synthesis and Future Research Agenda. *International Journal of Image and Graphics*, 2350061.
- [37] Al Mansoori, S., Almansoori, A., Alshamsi, M., Salloum, S. A., & Shaalan, K. (2020). Suspicious activity detection of Twitter and Facebook using sentimental analysis. *TEM Journal*, 9(4), 1313.
- [38] Alsaif, H. F., & Aldossari, H. D. (2023). Review of stance detection for rumor verification in social media. *Engineering Applications of Artificial Intelligence*, 119, 105801.
- [39] Guttikonda, J. B. (2019). A new steganalysis approach with an efficient feature selection and classification algorithms for identifying the stego images. *Multimedia Tools and Applications*, 78(15), 21113-21131.
- [40] Ghallab, A., Mohsen, A., & Ali, Y. (2020). Arabic sentiment analysis: A systematic literature review. *Applied Computational Intelligence and Soft Computing*, 2020, 1-21.
- [41] Ellaky, Z., Benabbou, F., & Ouahabi, S. (2023). Systematic Literature Review of Social Media Bots Detection Systems. *Journal of King Saud University-Computer and Information Sciences*.
- [42] Tursynova, A., & Omarov, B. (2021, November). 3D U-Net for brain stroke lesion segmentation on ISLES 2018 dataset. In *2021 16th International Conference on Electronics Computer and Computation (ICECCO)* (pp. 1-4). IEEE.

# Real-Time Road Surface Damage Detection Framework based on Mask R-CNN Model

Bakhytzhhan Kulambayev<sup>1</sup>, Magzat Nurlybek<sup>2</sup>, Gulnar Astaubayeva<sup>3</sup>, Gulnara Tleuberdiyeva<sup>4</sup>,  
Serik Zholdasbayev<sup>5</sup>, Abdimukhan Tolep<sup>6</sup>  
Turan University, Almaty, Kazakhstan<sup>1</sup>  
Bachelor Student at Turan University, Almaty, Kazakhstan<sup>2</sup>  
NARXOZ University, Almaty, Kazakhstan<sup>3,4</sup>  
International Information Technology University, Almaty, Kazakhstan<sup>5</sup>  
Khoja Akhmet Yassawi International Kazakh-Turkish University, Turkistan, Kazakhstan<sup>6</sup>

**Abstract**—In the ever-evolving realm of infrastructure management, the timely and accurate detection of road surface damages is imperative for the longevity and safety of transportation networks. This research paper introduces a pioneering framework centered on the Mask R-CNN (Region-based Convolutional Neural Networks) model for real-time road surface damage detection. The overarching methodology encapsulates a deep learning-based approach to discern and classify various road aberrations such as potholes, cracks, and rutting. The chosen Mask R-CNN architecture, renowned for its proficiency in instance segmentation tasks, has been fine-tuned and optimized specifically for the unique challenges posed by road surfaces under diverse lighting and environmental conditions. A diverse dataset, amalgamating urban, suburban, and rural roadways under varied climatic conditions, served as the foundation for model training and validation. Preliminary results have not only underscored the model's robustness in real-time detection but also its superiority in terms of accuracy and computational efficiency when juxtaposed with extant methods. Concomitantly, the framework emphasizes scalability and adaptability, positing it as a frontrunner for potential integration into automated road maintenance systems and vehicular navigation aids. This trailblazing endeavor elucidates the potentialities of deep learning paradigms in revolutionizing road management systems, thus fostering safer and more efficient transportation environments.

**Keywords**—Deep learning; CNN; random forest; SVM; neural network; prediction; analysis

## I. INTRODUCTION

Road infrastructure remains a pivotal element in the socio-economic fabric of nations, serving as the backbone of trade, transportation, and daily commuting [1]. As urbanization and globalization continue to expand, so does the reliance on a durable and well-maintained road network. While the necessity of pristine road infrastructure is universally recognized, it's equally undeniable that roadways are persistently subjected to degradation [2]. Factors such as climatic extremes, vehicular stress, and natural wear-and-tear all contribute to the deterioration of road surfaces [3]. The consequent damages, ranging from innocuous surface irregularities to perilous potholes, pose significant safety risks to motorists, exacerbate vehicular wear, and escalate maintenance costs. Hence, timely

and accurate damage detection is a sine qua non for effective road maintenance and ensuring commuter safety.

Historically, the task of road surface damage detection was primarily relegated to manual inspections. Field engineers and surveyors would periodically inspect stretches of road, logging visible damages for subsequent repair. However, such methods are inherently fraught with shortcomings. Human inspections are not only labor-intensive and time-consuming but are also marked by subjective biases and are often limited by the perceptual constraints of the human eye. Furthermore, large-scale road networks make manual monitoring a logistical challenge, often leading to significant delays between damage occurrence and its eventual rectification [4].

Emerging from this backdrop, technological solutions began to surface, attempting to alleviate the limitations of manual inspection. Early endeavors in this direction exploited image processing techniques to detect road anomalies [5]. While promising, these rudimentary techniques often grappled with issues of low accuracy, particularly in diverse environmental and lighting conditions [6]. More advanced techniques leveraging pattern recognition and machine learning offered an uptick in detection capabilities but remained hamstrung by their inability to perform adequately in real-time scenarios and their frequent misclassifications in complex road environments [7].

The recent upswing in the adoption of deep learning models across diverse domains signaled a transformative potential for road damage detection. Deep learning, a subset of machine learning, empowers models to learn and make decisions from vast amounts of data, often surpassing human-level performance in specific tasks [8]. In the context of road damage detection, Convolutional Neural Networks (CNNs) have emerged as a favored tool due to their adeptness in handling image data [9]. However, while CNNs are proficient in classification tasks, the intricate nature of road damage detection demands a more nuanced approach, one capable of instance segmentation—a task that goes beyond mere classification and seeks to delineate and identify specific objects within images.

This is where the Mask R-CNN model [10] enters the fray. An evolution of the established R-CNN [11] and Fast R-CNN [12] architectures, Mask R-CNN has proven its mettle in

instance segmentation tasks across various domains. Its unique architecture, which seamlessly integrates the strengths of both its predecessors, enables precise object localization and pixel-wise mask prediction. Such capabilities render it an intriguing prospect for the intricacies of road surface damage detection.

This research paper aims to exploit the prowess of the Mask R-CNN model in developing a comprehensive framework for real-time road surface damage detection. Drawing upon a meticulously curated dataset encompassing a myriad of road types and conditions, the study seeks to optimize and fine-tune the Mask R-CNN model for this specialized task. Moreover, this investigation delves deep into the challenges inherent in road damage detection, such as variable lighting, shadow effects, wet surfaces, and other environmental nuances. By addressing these complexities, the paper aims to elevate the discourse on automated road damage detection and present a robust, scalable, and efficient solution.

In doing so, this paper positions itself at the intersection of advanced deep learning paradigms and pressing infrastructural challenges. It aspires not just to contribute to academic discourse but also to catalyze tangible shifts in how road maintenance authorities across the globe approach the monumental task of road upkeep and safety assurance. By marrying the Mask R-CNN model's capabilities with the real-world demands of road damage detection, this study embarks on a journey to redefine the standards of road infrastructure management in the age of artificial intelligence.

## II. RELATED WORKS

The journey of automating road surface damage detection has been a progressive one, punctuated by incremental innovations and paradigm shifts. As this research navigates the waters of the Mask R-CNN model for real-time road surface damage detection, it is imperative to contextualize its approach within the broader framework of previous efforts in this domain. This section endeavors to provide a comprehensive review of related works, elucidating the trajectory of technological advances that have shaped the discourse on automated road damage detection.

### A. Traditional Image Processing Techniques

The inception of automated methodologies for road surface damage detection is deeply rooted in traditional image processing techniques. In the nascent stages, simple, yet effective algorithms such as edge detection, thresholding, and morphological operations were employed to discern road anomalies, primarily cracks and potholes. Pioneering research, exemplified by the work of [13], and made strides in this domain by harnessing wavelet transforms for enhanced crack detection. While these early techniques represented a significant leap from manual inspection, they were not without their limitations. Particularly, their susceptibility to variable environmental conditions, such as fluctuating lighting and shadows, frequently resulted in a high rate of false positives. Consequently, despite their foundational contributions, it became evident that more sophisticated approaches were needed to achieve the precision and reliability demanded by real-world applications in road maintenance.

### B. Machine Learning and Pattern Recognition

Transitioning from the foundational image processing methodologies, the domain witnessed a paradigm shift with the advent of machine learning and pattern recognition techniques. Here, the emphasis transitioned from raw image manipulation to extracting discernible features, which could then be classified using algorithms. A seminal contribution in this realm was made by [14], who adeptly combined texture-based feature extraction with Support Vector Machines (SVM) to pinpoint road cracks. This strategy elevated the accuracy of detection substantially. However, it also introduced the intricacy of manual feature engineering, a labor-intensive endeavor with potential for inconsistencies. Despite the undeniable advancement in damage detection these methods brought about, the challenges they posed emphasized the need for more automated and adaptive solutions, paving the way for the exploration of deep learning techniques in subsequent research.

### C. Deep Learning and CNNs

The renaissance of neural networks, especially Convolutional Neural Networks (CNNs), ushered in a new era for road damage detection. The beauty of CNNs lies in their ability to automatically learn features from raw image data without explicit manual feature engineering. Significant contributions in this realm include the work of [15], who developed a road damage detection and classification system based on deep CNNs. Their model was not only adept at identifying damages but also categorizing them into types like cracks, potholes, and patches. However, while CNNs were proficient in classifying damaged regions, delineating the exact boundaries of these damages remained a challenge.

### D. R-CNN and its Evolution

The introduction of Region-based Convolutional Neural Networks (R-CNN) signaled a quantum leap in object detection tasks. R-CNN and its evolutionary offshoots, Fast R-CNN and Faster R-CNN, integrated region proposal networks with CNNs, allowing precise object localization within images [16-18]. In the context of road damage detection, this meant an enhanced ability to identify and demarcate specific damaged regions within a broader road image. The works of [19] stand testament to the efficacy of Faster R-CNN in detecting and segmenting road damages.

### E. Instance Segmentation with Mask R-CNN

Delving deeper into the world of object detection, the Mask R-CNN model emerged as a revolutionary tool, bringing the nuance of instance segmentation to the fore. Building upon the foundation laid by its predecessors, the Faster R-CNN, the Mask R-CNN transcended mere object localization, offering pixel-wise mask prediction for each identified entity within an image [20]. This level of granularity made it an optimal candidate for tasks requiring meticulous delineation, such as road damage detection. Early explorations into the model's applicability, highlighted by studies like those of [21], exhibited promising outcomes. The ability of the Mask R-CNN to pinpoint and define road surface anomalies with precision underscored its potential to set a new benchmark in the domain, promising a convergence of accuracy and granularity hitherto unseen in earlier methodologies.

### F. Real-Time Detection Challenges

While the evolution of detection techniques marked notable advancements, the exigencies of real-time processing remained a pivotal concern. The operational demands of road maintenance necessitate not just accuracy, but also timeliness in damage detection. Architectures like YOLO (You Only Look Once) and SSD (Single Shot MultiBox Detector), as elucidated by researchers such as [22-23], respectively, heralded solutions emphasizing real-time object detection. Though not tailored explicitly for road anomalies, the underlying principles of these frameworks provide invaluable insights. They spotlight the intricate balance and potential trade-offs between detection speed and accuracy. Such considerations are paramount when envisioning a model that operates in dynamic real-world settings, reinforcing the need for an optimal blend of precision and promptness in any prospective road damage detection system.

### G. Adaptive Learning and Transfer Learning

With vast and diverse road networks, training models from scratch becomes computationally expensive. The concept of transfer learning, where models pre-trained on large datasets are fine-tuned for specific tasks, gained traction. The author in [24] explored transfer learning for road damage detection, leveraging models initially trained on datasets like ImageNet and adapting them to the specific nuances of road images.

In synthesizing the above, one discerns a clear trajectory: from basic image processing to the intricacies of deep learning, and from the broad strokes of object detection to the finesse of instance segmentation. This research positions itself at this evolving frontier, seeking to harness the potential of Mask R-CNN, not just in terms of detection accuracy but also in meeting the demands of real-time processing.

## III. MATERIALS AND METHODS

A comprehensive review of pertinent literature underscores the unparalleled efficacy of deep convolutional neural networks (DCNNs) in current scholarly investigations. A pivotal initial step involves segmenting roadway imagery to demarcate relevant classes, crucial for defect identification. Presently, CNN architectures, such as the SegNet [25] and U-Net [26], are gaining traction for their effectiveness in this domain. The challenge arises from the subtle grayscale variations in road imagery and the minimal contrast between the intended subject and its backdrop, compounded by incidental noise and unrelated elements. To navigate these challenges, a fully convolutional neural network (FCNN) employing an "encoder-decoder" configuration is utilized, yielding a binary output image [27]. The FCNN bifurcates into a convolutional segment—transforming the primary image into a feature-rich representation—and a segment producing the segmented output from these features. This architecture encompasses a series of convolutional strata, augmented by filters and subsequent sub-discretization tiers. By integrating upsampling with convolutional stages, the architecture reconstructs the initial image dimensions, subsequently crafting a likelihood matrix.

The CrackForest dataset comprises 117 snapshots, partitioned into training, testing, and validation subsets. For each image, 64x64 segments are extracted arbitrarily from both training and test sets. Image quality amplifies with gamma correction, enhancing neural network performance. A 95:5 ratio, emphasizing defects constituting at least 5% of the image, is deemed optimal. With 15,200 training fragments juxtaposed against 3,968 test fragments, the balance is deemed propitious for the deep learning process. The network's evaluation employs intersection over union metrics, complemented by binary similarity metrics. Weight initialization within FCNN layers leverages the Glorot technique, normalizing each layer's input distributions, thus mitigating internal covariance shifts. Optimization ensues via the Adam optimizer. Research concludes that an optimal 25-epoch training duration—split between an initial 5 epochs and a subsequent 20—is effective. Execution of the FCNN blueprint leverages both Keras and TensorFlow platforms. Upon training completion, the artificial neural network undergoes rigorous testing and validation using sample data.

In this study, an enhanced methodology trained existing Mask R-CNN models via TensorFlow's Object Detection API, aiming to augment road defect detection efficiency. These refined models subsequently underwent rigorous evaluations utilizing meticulously curated annotation datasets.

### H. Data Collection and Preparation

Traditionally, road surface damage detection relied on aerial images or imagery sourced from vehicle-mounted cameras. Aerial imaging poses practical challenges due to the intricacies involved in capturing such images, restricting its widespread application. Conversely, using imagery derived from vehicle-mounted cameras offers more pragmatic utility, considering the ease of data acquisition. This positions commonly available devices, like smartphones, as potential tools for damage detection, whether the processing occurs in situ or is offloaded to a remote server. Consequently, we developed a unique dataset encompassing six distinct categories of road damage, with each image meticulously annotated by hand.

Fig. 1 presents a visual guide to the diverse damage types, denoted by specific class names such as D20. The subsequent illustrative table segregates these damages into six primary categories, distinguishing between cracks and other deformities. Crack-based damages further bifurcate into linear and alligator cracks, while other categories span potholes, ruts, and anomalies like faded lane markings. Notably, the breadth of damage categories explored in our study outstrips the limited scopes of prior works. For instance, the approach proposed by [28] merely detects potholes under the D40 label, while Jana et al. [29] differentiates damages strictly as longitudinal or transverse. Further, preceding deep learning studies [30-33] primarily focus on identifying the mere presence or absence of damage.



Fig. 1. Road damage photos and classes for a model training.

### I. Annotation and Classification for Enhanced Damage Detection

To facilitate a refined categorization, our annotation data delineates 12 distinct classifications of road damage and associated features captured in the photographs. The Microsoft Visual Object Tagging Tool (VoTT) was instrumental in annotating these color images. Within each image, specifically its lower two-thirds, every discernible feature within our predefined classes was segmented and appropriately labeled. Table I elucidates the compiled annotation data.

Among these classifications, "Scratches on Markings" emerged as the most prevalent, boasting 3,360 segments. This was closely followed by "Linear Cracks" at 3,080 segments. On the rarer end, "Grid Cracks in Patchings" registered the least at 252 segments, succeeded by "Stains", "Manholes", and

"Potholes". For analytical rigor, the data segments were stratified into training, validation, and testing datasets at a proportion of 0.6:0.2:0.2, respectively.

In our comprehensive research, we established a refined taxonomy of annotation data that encompasses 12 unique classifications pertinent to road damage and its corresponding features as depicted in the photographic evidence. The intricacies of the annotation process were adeptly managed using the Microsoft Visual Object Tagging Tool (VoTT), which proved pivotal for effective categorization within the color images. A keen focus was directed towards the inferior two-thirds of each image. Within this portion, every feature that aligned with our pre-established categories was diligently segmented and given an appropriate label. For a detailed scholarly overview, readers are directed to Table I, which presents a thorough synthesis of the amassed annotation data.

TABLE I. ROAD IMAGES ANNOTATION DATA

Class ID	Classes	Training	Validation	Testing	Total
1	Linear crack	3080	660	660	4400
2	Grid crack	658	141	141	940
3	Pavement joins	854	183	183	1220
4	Patchings	448	96	96	640
5	Fillings	1344	288	288	1920
6	Pot-holes	406	87	87	580
7	Manholes	336	72	72	480
8	Stains	266	57	57	380
9	Shadow	1190	255	255	1700
10	Pavement markings	1414	303	303	2020
11	Scratches on markings	3360	720	720	4800
12	Grid crack in patchings	252	54	54	360
0	Total	13608	2916	2916	19440

Among these classifications, "Scratches on Markings" emerged as the most prevalent, boasting 3,360 segments. This was closely followed by "Linear Cracks" at 3,080 segments. On the rarer end, "Grid Cracks in Patchings" registered the least at 252 segments, succeeded by "Stains", "Manholes", and "Potholes". For analytical rigor, the data segments were stratified into training, validation, and testing datasets at a proportion of 0.6:0.2:0.2, respectively.

#### IV. PROPOSED NETWORK

In pursuit of an integrated solution for crack identification and their granular pixel-wise delineation, the contemporary Mask R-CNN convolutional network architecture was chosen. Delving into its foundation and operational mechanics, one finds that the Mask R-CNN is rooted in a lineage of convolutional neural networks designed for localized region processing. This lineage encompasses the Region-based Convolutional Neural Network (R-CNN), its subsequent iterations in Fast R-CNN, and the even more refined Faster R-CNN.

Fig. 2 portrays our adoption of the Mask R-CNN architecture tailored for road surface damage identification. At its core, the Mask R-CNN framework is intrinsically intricate in its block configuration. The initial phase involves the input image being processed through the network, highlighting a

feature map. Common feature extractors employed for this purpose include VGG-16, the 50-layer Residual Neural Network (ResNet50), and the more extensive 101-layer Residual Neural Network (ResNet101), with layers focused on classification being omitted. An evolutionary distinction of this architecture, setting it apart from earlier iterations, is the incorporation of the Feature Pyramid Network (FPN) methodology. This technique is pivotal in harvesting feature maps across varied scales. Within this paradigm, consecutive layers of the network, characterized by descending dimensions, are perceived as a stratified "pyramid", where lower tier maps are high-resolution, and the apex tiers possess enhanced semantic abstraction.

Post this feature map extraction, the Region Proposals Network (RPN) segment takes center stage. Its primary objective is to pinpoint hypothesized regions within the image that potentially harbor objects. This is achieved by sliding a 3x3 neural network window over the feature map, with the output anchored on predefined 'k anchors' – essentially frameworks with specified dimensions and orientations. For every such anchor, the RPN forecasts the object's presence and, if detected, fine-tunes the coordinates of the object's bounding box. This stage's ultimate goal revolves around spotlighting regions brimming with potential object presence. Consecutively, overlapping regions are eliminated, courtesy of the non-maximum suppression operation.

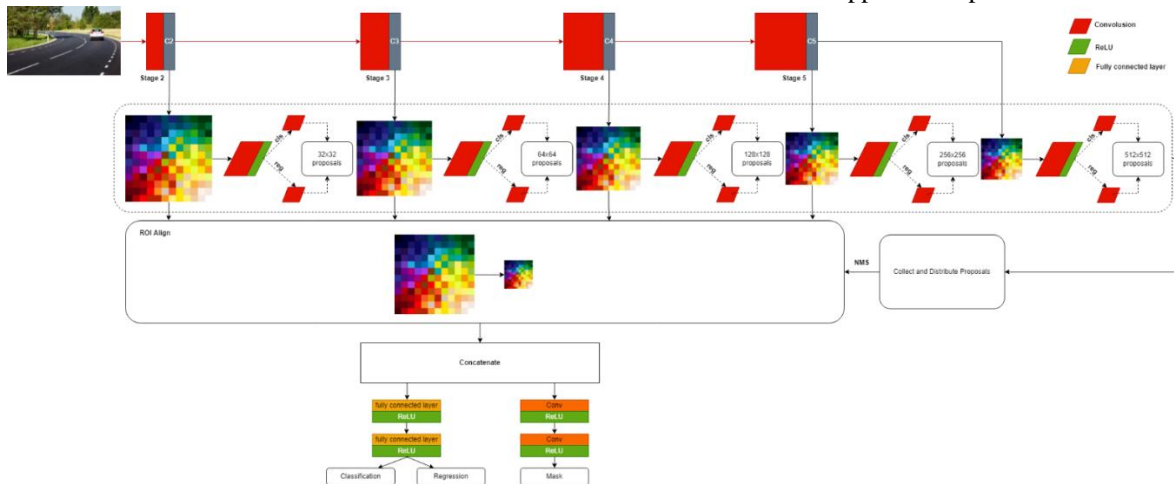


Fig. 2. Proposed mask R-CNN model for road surface damage detection.

In the subsequent phase, the Region of Interest (ROI) Align mechanism comes into play, selecting values pertinent to these regions from the feature maps and standardizing them to a uniform size. These harmonized values then undergo final processes including classification, adjustment of bounding box coordinates, and mask prediction. Notably, the emergent mask, despite its considerably diminished size, retains real values. Once the mask is scaled congruent to the object's dimensions, the precision achieved is commendable.

## V. EVALUATION

To ascertain the efficacy of the suggested model, it's evaluated against key metrics, specifically the mean average precision (MaP) and the average recall (AR), both scrutinized at varying thresholds of intersection over union (IoU). In scenarios involving classification paired with object localization and detection, the ratio derived from the areas of the bounding boxes frequently serves as a determinant metric, reflecting the accuracy of the bounding box placement.

Embedded within the Mask RCNN is a region proposal network layer, adept at executing parallel inferences concerning class categorization, segmentation, and mask territories, leading to a resultant of six distinctive loss metrics. Complementing these inherent model-specific metrics, both average precisions and average recalls, benchmarked at an IoU value of 0.5, are employed across all twelve delineated road object categories, as referenced in [34].

$$IoU = \frac{S(A \cap B)}{S(A \cup B)} \quad (1)$$

Given A as the forecasted bounding box and B as the reference bounding box, the Intersection over Union (IoU) serves as a metric. The IoU value stands at zero when the bounding boxes do not intersect, and reaches its zenith of one when the bounding boxes perfectly coincide.

A pivotal aim of evaluation is maximizing the detection of instances within a given population using a screening method. It's imperative that false negatives are curtailed, even if it necessitates an uptick in false positives. This emphasis necessitates the careful consideration of three fundamental metrics: the true positive rate (TPR), false positive rate (FPR), and overall accuracy (ACC). Within the realm of medical terminologies, TPR often finds its synonym in sensitivity (SEN) and is represented as seen in equation (2), as documented in [35].

$$TPR = SEN = \frac{TP}{P} \quad (2)$$

Let TP represent the count of true positives, while P signifies the total positive instances in the dataset.

The quantification of the subsequent metric, the false positive rate, is articulated in equation (3), as delineated in [36]:

$$FPR = \frac{FP}{N} \quad (3)$$

Where N denotes the aggregate count of negative cases in the population and FP symbolizes the number of false positives. Furthermore, the true negative instances are also represented by N. However, a more intuitive understanding of this metric is the fraction of true negatives out of the actual negative cases. In medical terminology, this metric is often referred to as specificity (SPEC), articulated as equation (4), as cited in [37]:

$$TNR = SPEC = \frac{TN}{N} = 1 - FPR \quad (4)$$

Ultimately, the metric of accuracy encapsulates the equilibrium between true positive and true negative outcomes. This metric becomes particularly insightful when there exists an imbalance between positive and negative instances within the dataset. This is quantitatively represented in equation (5), as referenced in [38]:

$$ACC = \frac{TP + TN}{P + N} \quad (5)$$

## VI. EXPERIMENTAL RESULTS

Within this segment, the experimental findings are bifurcated into two distinct subsections. The initial subsection elucidates the results pertaining to road damage detection, followed by an exposition on road damage segmentation outcomes. The subsequent section delves into the real-time performance of the proposed model, accompanied by visual demonstrations. This encompasses both original imagery and annotated representations of road conditions. In the third subsection, a comprehensive assessment of the model is presented, detailing evaluative metrics such as precision, recall, and F-score for the respective categories of road surface impairments.

### A. Road Damage Detection Results

Leveraging the intricacies of the Mask R-CNN architectural framework, we developed a nuanced system tailored for road damage detection. This state-of-the-art approach is adept at swiftly and accurately discerning multiple forms of roadway degradation, encompassing anomalies like cracks and spalling, as evidenced in the images procured using digital photographic equipment. To facilitate an insightful understanding and comparison of the system's performance, Table II meticulously catalogs the results of the damage detection endeavor. This tabulation emphasizes evaluative metrics, notably precision, recall, and the F1-score, underscoring the robustness and precision of the devised methodology.

### B. Road Damage Segmentation Results

In the process of isolating the segment of the image associated with the roadway, pixels within the road mask are accentuated. Subsequently, an 8-connected region search algorithm is employed on the resultant binary mask. The region boasting the highest pixel count is subsequently identified as the coverage mask, as depicted in a gray shade in Fig. 3.



TABLE II. EVALUATION OF THE PROPOSED METHOD BY CLASSES

Model	Precision	Recall	F1-score
<b>Proposed model</b>	<b>0.9214</b>	<b>0.9876</b>	<b>0.9571</b>
Fully convolutional encoder–decoder network [39]	0.9130	0.9410	0.9270
Deep learning-based semantic segmentation [40]	0.8340	0.6855	0.7524
UNet-based concrete crack detection CrackUnet19 [41]	0.9145	0.8867	0.9004
Two-step light gradient boosting machine [42]	0.6801	0.7578	0.6950
Semantic segmentation using deep learning [43]	0.4044	0.7847	0.4994
Automated vision-based detection [44]	0.9236	0.8928	0.9079



Fig. 3. Marked up road images.

To assess the proficiency of the devised methodology for defect detection, a curated dataset comprising 50 authentic images showcasing road cracks was meticulously assembled. Fig. 4 juxtaposes the outcomes of manual crack delineation against the segmentation outcomes achieved through the proposed neural network's granular pixel-wise selection on an actual image.

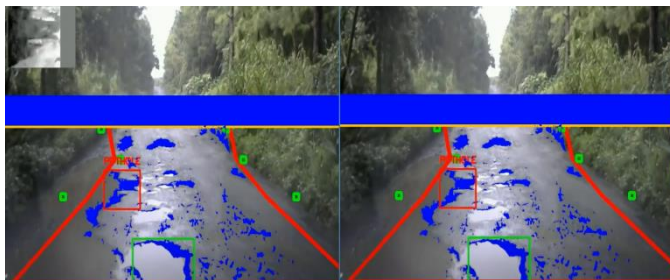


Fig. 4. Marked up pixel-wise selection of road images.

Fig. 5 presents the outcomes of model evaluation over 100 epochs. In Fig. 5, the accuracy and validation accuracy of the advanced model are delineated. It can be inferred from the data that our model achieves an approximate accuracy of 90% within 60 epochs, indicative of its robustness and applicability in real-world scenarios.

Fig. 6 depicts the training and validation loss associated with the model. The observed minimal loss suggests that the model is poised to commit minimal errors in practical applications.

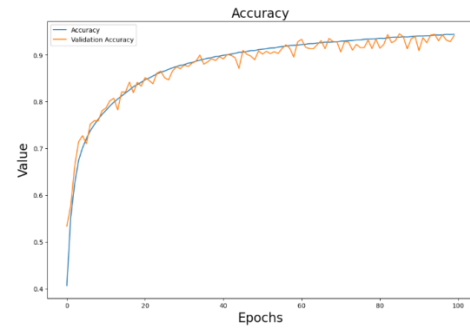


Fig. 5. Accuracy in road damage detection.

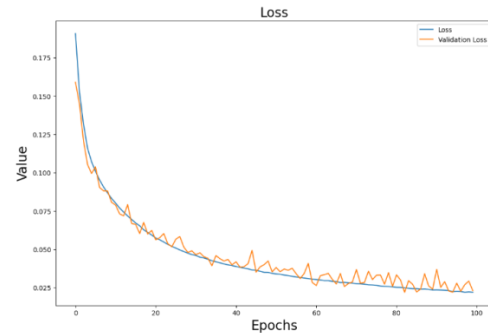


Fig. 6. Loss in road damage detection.

Various strategies employing deep learning paradigms aim to enhance road safety. Contemporary research offers innovative solutions to this issue [45]. For instance, [46] introduced a Vehicle Re-Identification technique to address challenges stemming from significant intra-class variances due to changing vehicle viewpoints during motion and pronounced inter-class resemblances due to analogous appearances. Our model is tailored to identify road surface imperfections using smartphone cameras or any equipment capable of capturing real-time road footage. Based on the results from the conducted experiments, it can be posited that deep learning techniques hold promise in addressing road safety and security challenges.

Table III presents the metrics associated with the model concerning bounding boxes and segmentation masks. For bounding boxes, the metrics for mAP at various IoU thresholds (IoU=.50:.05:.95), mAP (IoU=.50), and mAP (IoU=.75) register as 0.2432, 0.4382, and 0.2482, respectively. In contrast, these metrics for segmentation masks are discerned to be 0.1600, 0.3257, and 0.1279, marking a noticeable decline. The Precision mAP (small) for minuscule objects manifests as markedly lower values, being 0.0365 and 0.0133 for bounding boxes and segmentation masks, respectively, especially when juxtaposed against the Precision mAP for larger and medium-sized entities. The Average Recall metrics for small, medium, and large entities on bounding boxes are quantified as 0.1166, 0.3132, and 0.4717 respectively, whereas the corresponding values for segmentation masks are 0.1021, 0.2528, and 0.2732. Pertaining to our designated damage categories such as linear cracks (denoted as Crack1), grid cracks (labelled as Crack2), potholes, scratches on road markings, and grid cracks in surface repairs, the detection precision metrics at an IoU threshold of .50 are 0.4085, 0.4958, 0.5714, 0.5934, and 0.4000, respectively.

TABLE III. EVALUATION OF THE PROPOSED METHOD BY CLASSES

Classes	Precision @ 0.5 IoU (Bounding box)	Recall @ 0.5 IoU (Bounding box)	Recall @ 0.5 IoU (Segmentation)	Recall @ 0.5 IoU (Segmentation)
Linear crack	0.5383	0.3847	0.3583	0.2639
Grid crack	0.6256	0.7140	0.5920	0.6744
Pavement joins	0.4900	0.5179	0.2498	0.2531
Patchings	0.7644	0.5584	0.8161	0.5843
Fillings	0.6071	0.4667	0.3040	0.2528
Pot-holes	0.7012	0.4155	0.7012	0.4155
Manholes	0.9596	0.8798	0.9596	0.8798
Stains	0.1798	0.1484	0.1191	0.1282
Shadow	0.5273	0.4317	0.3285	0.2713
Pavement markings	0.7522	0.7460	0.5065	0.5002
Scratches on markings	0.7232	0.7531	0.4863	0.4944
Grid crack in patchings	0.5298	0.2474	0.7298	0.3063

## VII. CONCLUSION

## REFERENCES

This research delved deeply into the realm of road surface damage detection, harnessing the potential of the Mask R-CNN architecture. The imperative need to develop robust, accurate, and real-time systems for detecting and classifying road damages stems from the crucial role such systems play in ensuring roadway safety and aiding in timely maintenance. A cornerstone of infrastructure management, road health significantly impacts both economic metrics and public safety.

The Mask R-CNN model showcased its prowess in detecting various types of surface damages with commendable precision. Emphasis was placed on understanding its structural nuances and ensuring optimal parameter selection to refine the resultant models. Features like the Region Proposals Network and the integration of the Feature Pyramid Network brought depth and versatility to the proposed method, allowing it to contend with complex road scenarios.

Key metrics used in assessing the model, including mAP and Average Recall across varying IoU thresholds, offered insightful perspectives into the model's performance. The observed results were heartening, with the model showcasing proficiency, especially in differentiating between minor and significant road damage categories.

Comparative analyses with extant literature reinforced the efficacy of the proposed approach, especially considering the challenges posed by real-time, on-ground situations. The model's capacity to work with images and footage from commonplace devices, such as smartphones, stands testament to its applicability in real-world scenarios, democratizing road damage detection to a broader user base.

In summation, while the world of deep learning and neural networks continues to evolve, the application of these technologies in solving pertinent, real-world challenges, as showcased in this study, remains paramount. The presented work not only contributes a robust solution to road surface damage detection but also lays down a pathway for further refinement and innovation in the domain. As future directions, the integration of more advanced architectures and real-time response mechanisms can further elevate the impact and utility of such systems in global infrastructure management.

- [1] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler et al., "The cityscapes dataset for semantic urban scene understanding," In Proceedings of The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, Nevada, The US, pp. 3213-3223, 2016.
- [2] O. Zendel, K. Honauer, M. Murschitz, D. Steininger and G. Dominguez, "Wilddash-creating hazard-aware benchmarks," In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, pp. 402-416, 2018.
- [3] J. Zhang, Y. Sun, H. Liao, J. Zhu and Y. Zhang, "Automatic Parotid Gland Segmentation in MVCT Using Deep Convolutional Neural Networks," ACM Transactions on Computing for Healthcare, vol. 3, no. 2, pp. 1-15, 2021.
- [4] 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.
- [5] K. Gopalakrishnan, S. Khaitan, A. Choudhary and A. Agrawal, "Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection," Construction and building materials, vol. 157, no. 1, pp. 322-330, 2017.
- [6] Altayeva, A., Omarov, B., Suleimenov, Z., & Im Cho, Y. (2017, June). Application of multi-agent control systems in energy-efficient intelligent building. In 2017 Joint 17th World Congress of International Fuzzy Systems Association and 9th International Conference on Soft Computing and Intelligent Systems (IFSAS-SCIS) (pp. 1-5). IEEE.
- [7] S. Wu, J. Fang, X. Zheng and X. Li, "Sample and structure-guided network for road crack detection," IEEE Access, vol. 7, no. 1, pp. 130032-130043, 2019.
- [8] M. Maniat, C. Camp and A. Kashani, "Deep learning-based visual crack detection using Google Street View images," Neural Computing and Applications, vol. 33, no. 21, pp. 14565-14582, 2021.
- [9] D. Dewangan and S. Sahu, "RCNet: road classification convolutional neural networks for intelligent vehicle system," Intelligent Service Robotics, vol. 14, no. 2, pp. 199-214, 2021.
- [10] M. Masud, M. Hossain, H. Alhumyani, S. Alshamrani, O. Cheikhrouhou et al., "Pre-trained convolutional neural networks for breast cancer detection using ultrasound images," ACM Transactions on Internet Technology, vol. 21, no. 4, pp. 1-17, 2021.
- [11] S. Bang, S. Park, H. Kim, Y. Yoon and H. Kim, "A deep residual network with transfer learning for pixel-level road crack detection," Network, vol. 93, no. 84, pp. 89-03, 2018.
- [12] Y. Chen, H. Wang, W. Li, C. Sakaridis, D. Dai et al., "Scale-aware domain adaptive faster r-cnn," International Journal of Computer Vision, vol. 129, no. 7, pp. 2223-2243, 2021.
- [13] D. Quang and S. Bae. "A hybrid deep convolutional neural network approach for predicting the traffic congestion index," Promet-Traffic & Transportation, vol. 33, no. 3, pp. 373-385, 2021.

- [14] N. Safaei, O. Smadi, B. Safaei and A. Masoud, "Efficient road crack detection based on an adaptive pixel-level segmentation algorithm," *Transportation Research Record*, vol. 2675, no. 9, pp. 370-381, 2021.
- [15] S. Bang, S. Park, S., Kim and H. Kim, "Encoder-decoder network for pixel - level road crack detection in black - box images," *Computer - Aided Civil and Infrastructure Engineering*, vol. 34, no. 8, pp. 713-727, 2019.
- [16] V. Tran, T. Tran, H. Lee, K. Kim, J. Baek et al., "One stage detector (RetinaNet)-based crack detection for asphalt pavements considering pavement distresses and surface objects," *Journal of Civil Structural Health Monitoring*, vol. 11, no. 1, pp. 205-222, 2021.
- [17] Z. Lingxin, S. Junkai and Z. Baijie, "A review of the research and application of deep learning-based computer vision in structural damage detection," *Earthquake Engineering and Engineering Vibration*, vol. 21, no. 1, pp. 1-21, 2022.
- [18] K. Gopalakrishnan, S. Khaitan, A. Choudhary and A. Agrawal, "Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection," *Construction and Building Materials*, vol. 157, no. 1, pp.322-330, 2017.
- [19] S. Patra, A. Midya and S. Roy, "PotSpot: Participatory sensing based monitoring system for pothole detection using deep learning," *Multimedia Tools and Applications*, vol. 80, no. 16, pp. 25171-25195, 2021.
- [20] T. Rateke and A. Von Wangenheim, "Road surface detection and differentiation considering surface damages," *Autonomous Robots*, vol. 45, no. 2, pp. 299-312, 2021.
- [21] F. Yang, L. Zhang, S. Yu, D. Prokhorov, X. Mei et. al, "Feature pyramid and hierarchical boosting network for pavement crack detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 4, pp. 1525-1535, 2019.
- [22] Q. Zou, Y. Cao, Q. Li, Q. Mao and S. Wang, "CrackTree: Automatic crack detection from pavement images," *Pattern Recognition Letters*, vol. 33, no. 3, pp. 227-238, 2012.
- [23] Y. Shi, L. Cui Z. Qi, F. Meng and Z. Chen, "Automatic road crack detection using random structured forests," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 12, pp. 3434-3445, 2016.
- [24] H. Maeda, Y. Sekimoto, T. Seto, T. Kashiyama and H. Omata, "Road damage detection and classification using deep neural networks with smartphone images," *Computer - Aided Civil and Infrastructure Engineering*, vol. 33, no. 12, pp. 1127-1141, 2018.
- [25] H. Afify, K. Mohammed and A. Hassanien, "An improved framework for polyp image segmentation based on SegNet architecture," *International Journal of Imaging Systems and Technology*, vol. 31, no. 3, pp. 1741-1751, 2021.
- [26] B. Omarov, A. Tursynova, O. Postolache, K. Gamry, A. Batyrbekov et al., "Modified UNet Model for Brain Stroke Lesion Segmentation on Computed Tomography Images," *CMC-Computers, Materials & Continua*, vol. 71, no. 3, pp. 4701-4717, 2022.
- [27] D. Laredo, S. Ma, G. Leylaz, O. Schütze and J. Sun, "Automatic model selection for fully connected neural networks," *International Journal of Dynamics and Control*, vol. 8, no. 4, pp. 1063-1079, 2020.
- [28] H. Maeda, T. Kashiyama, Y. Sekimoto, T. Seto and H. Omata, "Generative adversarial network for road damage detection," *Computer - Aided Civil and Infrastructure Engineering*, vol. 36, no. 1, pp. 47-60, 2020.
- [29] S. Jana, S. Thangam, A. Kishore, V. Sai Kumar and S. Vandana, "Transfer learning based deep convolutional neural network model for pavement crack detection from images," *International Journal of Nonlinear Analysis and Applications*, vol 13, no. 1, pp. 1209-1223, 2022.
- [30] B. Kim, N. Yuvaraj, K. Sri Preethaa and R. Arun Pandian, "Surface crack detection using deep learning with shallow CNN architecture for enhanced computation," *Neural Computing and Applications*, vol. 33, no. 15, pp. 9289-9305, 2021.
- [31] Kulambayev, B., Beissenova, G., Katayev, N., Abduraimova, B., Zhaidakbayeva, L., Sarbassova, A., ... & Shyrakbayev, A. (2022). A Deep Learning-Based Approach for Road Surface Damage Detection. *Computers, Materials & Continua*, 73(2).
- [32] E. Protopapadakis, A. Voulodimos, A. Doulamis, N. Doulamis and T. Stathaki, "Automatic crack detection for tunnel inspection using deep learning and heuristic image post-processing," *Applied intelligence*, vol. 49, no. 7, pp. 2793-2806, 2019.
- [33] Jayakumar, L., Chitra, R. J., Sivasankari, J., Vidhya, S., Alimzhanova, L., Kazbekova, G., ... & Teressa, D. M. (2022). QoS Analysis for Cloud-Based IoT Data Using Multicriteria-Based Optimization Approach. *Computational Intelligence and Neuroscience*, 2022.
- [34] D. Russo, K. Zorn, A. Clark, H. Zhu and S. Ekins, "Comparing multiple machine learning algorithms and metrics for estrogen receptor binding prediction," *Molecular pharmaceutics*, vol. 15, no. 10, pp. 4361-4370, 2018.
- [35] V. Thambawita, D. Jha, H. Hammer, H. Johansen, D. Johansen et al., "An extensive study on cross-dataset bias and evaluation metrics interpretation for machine learning applied to gastrointestinal tract abnormality classification," *ACM Transactions on Computing for Healthcare*, vol. 1, no. 3, pp. 1-29, 2020.
- [36] 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.
- [37] Sultan, D., Omarov, B., Kozhamkulova, Z., Kazbekova, G., Alimzhanova, L., Dautbayeva, A., ... & Abdrakhmanov, R. (2023). A Review of Machine Learning Techniques in Cyberbullying Detection. *Computers, Materials & Continua*, 74(3).
- [38] S. Guillon, F. Joncour, P. Barrallon and L. Castanié, "Ground-truth uncertainty-aware metrics for machine learning applications on seismic image interpretation: Application to faults and horizon extraction," *The Leading Edge*, vol. 39, no. 10, pp. 734-741, 2020.
- [39] M. Islam and J. Kim, "Vision-based autonomous crack detection of concrete structures using a fully convolutional encoder-decoder network," *Sensors*, vol. 19, no. 19, pp. 4251, 2019.
- [40] T. Yamane and P. Chun, "Crack detection from a concrete surface image based on semantic segmentation using deep learning," *Journal of Advanced Concrete Technology*, vol. 18, no. 9, pp. 493-504, 2020.
- [41] L. Zhang, J. Shen and B. Zhu, "A research on an improved Unet-based concrete crack detection algorithm," *Structural Health Monitoring*, vol. 20, no. 4, pp. 1864-1879, 2021.
- [42] P. Chun, S. Izumi and T. Yamane, "Automatic detection method of cracks from concrete surface imagery using two - step light gradient boosting machine," *Computer - Aided Civil and Infrastructure Engineering*, vol. 36, no. 1, pp. 61-72, 2021.
- [43] D. Lee, J. Kim and D. Lee, "Robust concrete crack detection using deep learning-based semantic segmentation," *International Journal of Aeronautical and Space Sciences*, vol. 20, no. 1, pp. 287-299, 2019.
- [44] B. Kim and S. Cho, "Automated vision-based detection of cracks on concrete surfaces using a deep learning technique," *Sensors*, vol. 18, no. 10, pp. 3452, 2018.
- [45] Omarov, B., Narynov, S., Zhumanov, Z., Gumar, A., & Khassanova, M. (2022). A Skeleton-based Approach for Campus Violence Detection. *Computers, Materials & Continua*, 72(1).
- [46] X. R. Zhang, X. Chen, W. Sun, X. Z. He, "Vehicle Re-Identification Model Based on Optimized DenseNet121 with Joint Loss", *Computers, Materials & Continua*, vol. 67, no. 3, pp. 3933-3948, 2021.

# Osteoporosis Detection and Classification of Femur X-ray Images Through Spectral Domain Analysis using Texture Features

Dhanyavathi A, Veena M B

Department of Electronics and Communications, BMS College of Engineering, Bengaluru, India  
Visvesvaraya Technological University, Belagavi-590018, India

**Abstract**—Osteoporosis commonly diagnosed as a bone disorder that affects the significant portion of the population. The Dual X-ray Absorptiometry (DXA) is one of the most accepted standard methods of analyzing the bone disorder, but it is exorbitant. However X-ray is a cost effective, therefore the proposed work introduces a new technique to improve osteoporosis detection and classification of femur bone X-ray image. The spectral based sub band images texture features are used to analyze the Region Of Interest (ROI) femoral head trabecular bone. A spectral domain based on the Two-Dimensional Discrete Wavelet Transform (2D-DWT) is used to represent variations in finer details in the image. Trabecular femur bone texture is determined only by horizontal, vertical, and diagonal sub bands of DWT coefficients. The sub band images are further enhanced by applying the maximum response filter (MRF) at different scales, thereby enhancing the most significant responses. Consequently, the sum of the MRFs of different scale images is considered as the supervised database. To detect osteoporosis, the test and supervised images are analyzed to calculate two significant attributes such as Zero Mean Normalized Cross-Correlation (ZMNC) and Sum Squared Difference (SSD). Based on experimental results, the performance metrics measure is improved in all aspects over current methods.

**Keywords**—Classification; feature; femur; images; normal; osteopenia; osteoporosis; texture

## I. INTRODUCTION

This The disease osteoporosis causes loss of bone density and increases the risk of fractures in millions of people worldwide [1][2]. Debilitating fractures can be effectively managed and prevented with early detection. Osteoporosis is often diagnosed with X-ray imaging, but traditional methods often rely on visual assessment, which may be subjective and subject to human errors [3]. A Convolution Neural Network (CNN) model was used to assess osteoporosis based on hip radiographs. An ensemble model with clinical covariates was also investigated [4]. From a single Dual-Energy X-Ray Absorptiometry (DXA) image of the proximal femur, reconstruct both the 3D bone shape and the Three Dimension Bone Mass Density (3D-BMD) distribution [5]. A set of Quantitative Computed Tomography (QCT) scans, a statistical model of the combined shape and BMD distribution is constructed to detect osteoporosis [6]. A method of estimating the apparent physical BMD of the proximal femur from CT images with good accuracy when evaluating post-menopausal

osteoporosis. The proximal femur radiographs were analyzed using Gabor filters, wavelet transformations, and fractal dimensions-based texture analysis methods to identify osteoporosis [7]. The volumetric estimation of femur bone based on an x-ray image using a computer-based algorithm to detect osteoporosis [8]. A comparison of BMD of CT scan and BMD revealed a difference of 4.53 percent in volume. An analysis was conducted to examine how they related with BMD and anthropometric factors such as height and weight [9]. In recent study, 34% of Indian women had osteoporosis and 20% had osteopenia, respectively. The study measured the energy of the proximal femur trabecular bone as a result of osteoporosis postmenopause using dual-tree complex wavelet transforms (DT-CWT) [10]. DT-CWT has been successfully used to analyze the trabecular pattern on the right proximal femur on radiographs. The Gabor filter was used to calculate features from the trabecular pattern recorded on proximal femur radiographs in the assessment of osteoporosis [11]. In order to justify the classification result, Singh indexes of trabecular pattern are used. An expert system designed for diagnosing osteoporosis based on measuring bone texture using fuzzy X-ray images [12]. A fuzzy X-ray imaging technique analyzes trabecular bone texture and thus calculates bone density by combining resolution enhancement algorithms and edge detection algorithms. Both algorithms are efficient at calculating disease severity. An image of a femur bone can be classified using morphometric features from the image segmented [13]. The femur bone structure is segmented using active contour method from a 2D X-ray radiograph and morphometric measurements are calculated in pixel values for head diameter, head height, neck diameter, and intertrochanteric distance in the proximal femur and neck. Several images of patients with osteoporotic, osteopenia, and normal conditions are collected using computer tomography (CT) [14]. MIMICS software is used to analyze the condition more effectively. Initially, the images must be imported into the MIMICS software for analysis. An effected, normal femur bone is generated in 3D by MIMICS software and osteopenia and normal patients can use the analysis as a precaution. Gradient Harmony Search (GHS) optimization based deep networks are used for classification [15]. A Harmonic Search (HS) algorithm is incorporated with a Gradient Descent (GD) algorithm is used to build GHS. Utilizing machine learning and image processing techniques to detect early-stage fractures caused by osteoporosis in femur image [16]. The Fracture Risk Assessment tool (FRAX) calculations were performed

retrospectively on 560 volunteers (age at least 50 years) who underwent hip-spine X-rays, BMD scans and FRAX tool calculations [17]. To determine whether Cortical Thickness Index (CTI) and Canal Flare Index (CFI) are used to calculate neck BMD (nBMD) on anteroposterior radiographs, both indices were measured on anteroposterior radiographs. Based on characterization of BMD and trabecular bone microarchitecture, an automated approach was developed to predict biomechanical bone strength in proximal femur specimens [18]. As a diagnostic biomarker for osteoporosis diagnosis, tracking disease progression, and evaluating the response to therapeutic intervention, the automated and objective way trabecular bone microarchitecture is analyzed, and the subsequent reduction performance achieved suggest that it may be utilized in this manner. Two Deep learning Convolution Neural Networks (DCNNs) i.e., Alex Net and Google Net were trained on anteroposterior hip radiograph images to detect risk in the neck of femur [19]. Feasible element analysis (FEA) of the hip guided by high-resolution magnetic resonance imaging (MRI) has been developed to assess subject-specific bone strength [20]. If technique is further validated, management of hip fracture risks in the clinic may be useful. Researchers evaluated the involvement of spinal and hip flexion discordances in Korean patients with a typical femoral fractures and femur neck fractures [21]. Discordances might be affected by osteoporotic fracture locations. Using conventional radiographs obtained for various indications, a robust opportunistic screening tool for osteoporosis and fracture risk assessment was demonstrated to provide vertebral compression fractures detection, BMD estimation, and fracture risk estimation in a fully automated manner [22]. In assessing fracture risk, parameters and their interactions are analyzed [23]. In a multiple regression analysis, bone density and the loading directions in a sideways fall have been considered as independent variables along with the fracture risk index as a dependent variable. As independent variables, angle about the femoral neck axis at the coronal and transverse ends of the shaft was measured in both coronal and transverse planes. In the interaction analysis of parameters, bone density appears to have a greater effect on fracture risk. Although analyzing the current techniques for detecting osteoporosis in femur images helps to a definite conclusion, while several challenges remain, there is a need for the work to be advanced. To produce more comprehensive work and enhance system performance, a new approach of osteoporosis detection in femur image based on spectrum analysis is introduced in this work.

## II. METHODOLOGY FOR PROPOSED WORK

The Fig. 1 illustrates the proposed work analysis is on the basis of spectral domain, 2D-DWT spectral domain analysis is introduced to analyze texture features of the X-ray of ROI femur images, offering valuable insights beyond the visual representation.

An analysis of the spectral features of image texture in two dimensions is carried out by one level decomposition of 2D-DWT [24][25]. Detecting osteoporosis involves two steps: the first is analyzing texture features on ROI images of femur bones and then deciding whether the given test image is normal (healthy bone) or abnormal (osteopenia or osteoporosis). By focusing on Horizontal Coefficients (HC), Vertical

Coefficients (VC), and Diagonal Coefficients (DC), the feature dimension in 2D-DWT can be reduced substantially. As a result, the proposed system model can detect bone diseases effectively with this reduced set of image texture. Ultimately, this work aims to obtain meaningful information from texture features by using symmetric wavelet family. With wavelet transformations, extracting texture information at different directions depending on how it is oriented. Further the image texture analysis of 2D-DWT is executed using MRFs at different scales. In order to classify normal or abnormal images, test and supervised images are matched based on attributes of ZMNCC and SSD of MRFs.

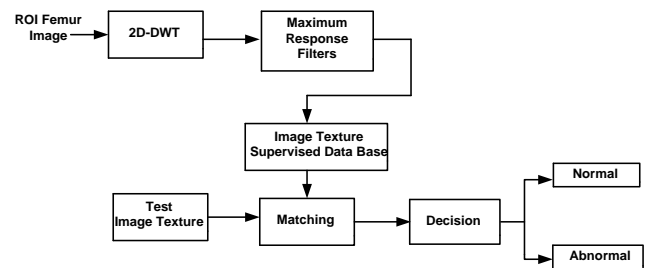


Fig. 1. Block schematic for the suggested work.

### A. Two Dimension Discrete Wavelet Transform

The 2D-DWT is a versatile and powerful tool for analyzing and processing 2D data like images, and its significance be situated in its capability to provide a compact and meaningful representation that facilitates various texture analysis in spectral domain. There were four sub band images in a one level decomposed 2D-DWT is illustrated in Fig. 2. Rows of the input image matrix are first transformed with an LPF and HPF, then its columns are treated with an LPF and HPF to produce sub band images are like approximation coefficients i.e., AC and three detail sub band images like HC, VC, and DC correspond to horizontal, vertical, and diagonal coefficients, respectively and each sub band images size are quarter of the input image. All three detail coefficients are represented half-resolutions of the input image and edge variations are almost exclusively visible in these three images. These three sub band images are considered in this study because of their texture information gives significant variations of intensity for analyzing the trabecular micro architecture of the femur bone.

The Symlet-4 wavelet is the basis function used in this work, due to its higher-scale detail coefficients are captured as finer details and texture patterns, while lower-scale detail coefficients capture broader texture patterns. The Symlet-4 is a type of wavelet that is used due to its balance between compact support and frequency localization. The sym in Sym4 stands for symmetric, which means that it has a symmetric shape. It is like Daubechies wavelets but have slightly different properties. The significance of using the Symlet-4 wavelet for image texture analysis lies in its ability to capture both low-frequency and high-frequency information effectively.

### B. Two-Dimensional Maximum Response Filter

A two-dimensional maximum response filter (2D-MRF) is introduced in this work to extract fine and coarse detail information of HC, VC and DC and its kernel is based on a Gaussian filters [26][27] at different scales. Each pixel location

must be analyzed in order to obtain the response values, first convolve the image with the Gaussian filter at multiple scales. Thus, the filtered output is calculated by taking the maximum value from the responses for each pixel location.

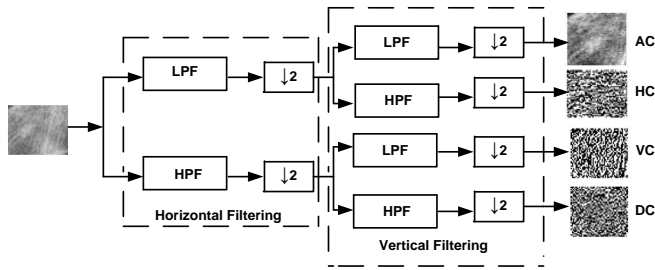


Fig. 2. First level 2D-DWT decomposition structure.

The MRF can help improve the robustness of feature extraction methods to noise and variations in the image. By emphasizing the most prominent responses and suppressing less significant ones, the filter helps in reducing the influence of noise or unwanted artifacts on the extracted features. The MRF can exhibit scale properties i.e., it can be effective in detecting the same pattern at different scales within the image. It helps in reducing the computational load by focusing on the most relevant features, this results in a faster processing and analysis of image datasets. Eq. (1) is the 2D-MRF using a Gaussian function, given an input sub band images  $S_B(x, y)$ , which are HC, VC and DC.

$$G(x, y) = \frac{1}{(2\pi\sigma^2)} e^{-\frac{(x^2+y^2)}{(2\sigma^2)}} \quad (1)$$

However, the Gaussian filter is characterized by a standard deviation, which is a scaling factor of the filter to compute the response values at each pixel location in the image i.e., where is the different scales and perform convolution operation ( $*$ ) between the image and the Gaussian filter at multiple scales shown in (2).

$$R_V(x, y, S) = S_B(x, y) * G(x, y, S) \quad (2)$$

Each pixel location is evaluated at multiple scales after obtaining the response values, a maximum response filter selects the response value that is the highest i.e., from the set of responses for each pixel demonstrated in (3).

$$R_{max}(x, y) = \max(R_V(x, y, 1), R_V(x, y, 2), \dots, R_V(x, y, S)) \quad (3)$$

The total number of scales used in this work are two different sets i.e., [0.3, 0.6, 0.9] and [5, 10, 15] to get fine and coarse detailed texture information respectively. Finally, the filtered output  $F_o(x, y)$  as in (4) represents the maximum pixel value at each location.

$$F_o(x, y) = R_{max}(x, y) \quad (4)$$

The MRFs process efficiently highlights the most salient features across different scales and emphasizes significant structures while suppressing less important ones. The scale

values determine the filter output i.e., Gaussian filter standard deviation ( $\sigma$ ) can be adaptable to for fine and coarse image texture requirements. The total texture image  $T_I(x, y)$  is a supervised image texture data in this proposed work, its value is calculated at each pixel location  $(x, y)$  in the image using the following algorithm.

**Algorithm :** Determination of fine and coarse texture image information

**Step1:** Fine texture analysis

$$\text{Find, } F_1(x, y) = R_{max}(x, y)$$

$$R_{max}(x, y) = \max(R_H(x, y, 0.3), R_V(x, y, 0.3), R_D(x, y, 0.3))$$

The  $R_H(x, y, 0.3)$  is filter response due to HC input at 0.3 scale, the  $R_V(x, y, 0.3)$  is filter response due to VC input at 0.3 scale and  $R_D(x, y, 0.3)$  is filter response due to DC input at 0.3 scale.

Similarly, find  $F_2(x, y) = R_{max}(x, y)$  at 0.6 scale

and  $F_3(x, y) = R_{max}(x, y)$  at 0.9 scale.

**Step 2:** Determine fine texture sum i.e.,  $S_F(x, y)$

**Step 3:** Coarse texture analysis

Repeat Step 1: to find  $F_4(x, y)$  at 5 scale,

$F_5(x, y)$  at 10 scale and  $F_6(x, y)$  at 15 scale.

**Step 4:** Determine coarse texture sum  $S_C(x, y)$

$$S_C(x, y) = F_4(x, y) + F_5(x, y) + F_6(x, y)$$

**Step 5:** Determine total texture image i.e.,  $T_I(x, y)$

$$T_I(x, y) = S_F(x, y) + S_C(x, y)$$

**C. Matching and Decision process**

A general match is based on finding test features that correspond to supervised databases [28]. In the matching process in which the test images and supervised images texture features are matching based on two attributes such as Zero Mean Normalized Cross-Correlation (ZNCC) and Sum of Squared Difference (SSD), which are obtained from total texture image i.e., The classification of the input test image is developed on having the highest possible value of ZNCC and lowest value of SSD, which are considered as the matching test classes. Pixels are compared based on their intensity values. These attributes help quantify the similarity between two images. Based on their intensity values, these attributes help quantify the similarity between two images.

- Zero Mean Normalized Cross-Correlation (ZNCC): Using the ZNCC, the cross-correlation between two images can be normalized. An image's similarity can be measured with it, where a value close to 1 indicates high similarity, while a value close to -1 indicates high dissimilarity.

Eq. (5) is the ZNCC between two images supervised i.e. test given by:

$$ZNCC(S_I, T) = \frac{\sum [(S_I(i, j) - \mu_{(S_I)}) (T(i, j) - \mu_{(T)})]}{\sqrt{\sum (S_I(i, j) - \mu_{(S_I)})^2} \sqrt{\sum (T(i, j) - \mu_{(T)})^2}} \quad (5)$$

Where:  $S_I(i, j)$  is the intensity value of supervised image  $S_I$  at position  $(i, j)$ . The  $T(i, j)$  is the intensity value of test image  $T$  at position  $(i, j)$ , The  $\mu_{(S_I)}$  is the mean intensity value of image. The is the mean intensity value of image  $T$ , the  $\sum$  denotes the summation over all pixel positions  $(i, j)$  in the images.

- Sum of Squared Difference (SSD): It measures the difference between corresponding intensity of the supervised image and the test image by summarizing their squares. Using the (6), it measures the degree of two images diverge from one another.

$$SSD(S_I, T) = \sum (S_I(i, j) - T(i, j))^2 \quad (6)$$

where, the summation is performed over all pixel positions in the images. In SSD case, test and supervised images are more similar when the value is lower. Based on SSD and ZNCC, the test input ROI image class is identified by reflecting on  $D1 = \text{Max}\{ZCCC\}$  and  $D2 = \text{min}\{SSD\}$ , if both D1 and D2 conditions are satisfied, that class is reflected as prediction class. The three classes are employed in this work as abnormal labeled as class-1 (osteopenia), class-2 (osteoporosis) and class-3 (normal).

### III. PERFORMANCE METRICS

The confusion matrix provides valuable insights into the model performance by quantifying different types of classifications [29]. The Table I displays the general confusion matrix used in the field of medicine to categorize the patient's positive and negative health conditions. From the confusion matrix, calculate various evaluation metrics, such as accuracy, precision, recall, and F1 score, which help assess the model effectiveness in different aspects.

TABLE I. GENERAL REPRESENTATION OF CONFUSION MATRIX

	Predicted Negative	Predicted Positive
Actual Negative	TN (True Negative)	FP (False Positive)
Actual Positive	FN (False Negative)	TP (True Positive)

- 1) *True positive (TP)*: Indicates that, number of samples were positive predictions made correctly.
- 2) *True negative (TN)*: Number of negatively predicted samples that were correct.
- 3) *False positive (FP)*: Incorrectly predicted positive samples are measured by this metric.
- 4) *False negative (FN)*: Number of negative samples predicted incorrectly.

In accordance with the confusion matrix, here are various evaluation metrics: Model accuracy measures how accurate the predictions as in (7),

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (7)$$

Precision (Positive Predictive Value, PPV), estimates the positive samples predicted correctly by the model using (8),

$$Precision = \frac{TP}{(TP + FP)} \quad (8)$$

Recall (Sensitivity, True Positive Rate, TPR) measures the model ability to correctly identify positive samples among all actual positive samples as in (9),

$$Recall = \frac{TP}{(TP + FN)} \quad (9)$$

Specificity (True Negative Rate, TNR) measures the model's ability to correctly identify negative samples among all actual negative samples measured in (10),

$$Specificity = \frac{TN}{(TN + FP)} \quad (10)$$

The F1 score represents in (11) is a balanced measure of the model's performance by combining precision and recall together.

$$F1\ score = \frac{2(Precision * Recall)}{(Precision + Recall)} \quad (11)$$

Analyzing the performance of a model using these metrics provides valuable insight into different aspects. A high accuracy indicates overall good performance, it indicates good class predictions when precision and recall are high. As it incorporates both precision and recall, the F1 score is particularly useful when both are equally important. Interpretation of these evaluation metrics must consider the context. If there is an imbalanced dataset, accuracy might not be a reliable measure, and other metrics like precision-recall curve or area under the receiver operating characteristic (AU-ROC-) curve might be more informative [30]. This curve plots between two parameters: True Positive Rate (TPR), False Positive Rate (FPR).

### IV. DATASET DESCRIPTION

Total 51 X-ray femur images were collected with the focal distance was set at 0.812 m. The X-ray parameters were 75-80 kV and 80 mAs for all patients. This study used images supplied by a reputed Bangalore hospital, Karnataka state, India. The image data consist of 2D radiographic images in JPEG format of size 2140×1760. In that 23 are normal, 10 are osteopenia and 18 are osteoporosis femur images. Table II listed the region of interested (ROI) femur bone of left and right of size 170×114, the total 102 ROI images including both left and right, which are considered to verify the proposed system experimentally.

TABLE II. DESCRIPTION OF DATA SET

Total ROI images: (23×2) + (10×2) + (18×2)=102					
Normal femur bone		Abnormal femur bone			
		Osteopenia femur		Osteoporosis femur	
Left	Right	Left	Right	Left	Right
23	23	10	10	18	18

A. Experimental Dataset Division

The proposed model can be experimentally tested by dividing the 102 total ROI of the image dataset into different cross-folding schemes, in the following Table III. A comparison between this and a previously conducted study on cross-folding methods reveals how well the proposed model performs.

TABLE III. EXPERIMENTAL DIVISION OF DATASET

Data - folding	Total number of samples of ROI images = 102					
	Supervised set			Test set		
	Normal	Abnormal		Normal	Abnormal	
	Osteopenia	Osteoporosis		Osteopenia	Osteoporosis	
Two-fold	12+12	5+5	9+9	11+11	5+5	9+9
	24+10+18=52			22+10+18=50		
Three-fold	16+16	7+7	12+12	7+7	3+3	6+6
	32+14+24=70			14+6+12=32		
Four-fold	18+18	8+8	14+14	5+5	2+2	4+4
	36+16+28=80			10+4+8=22		

V. RESULTS AND DISCUSSION

The input image of femur is a very low-quality image because the X-ray image of inner trabecular bones micro-architecture is not visible to distinguish between healthy bone and osteoporotic bone. Because of its versatility, DWT has become a very useful method for the process of decomposing an image into several resolutions through wavelet decomposition. By concentrating the wavelet energy in time and keeping its periodic properties, wavelets can simultaneously analyze both time and frequency of pixels intensity in the image. By decomposing a digital image into different sub bands, the 2D-DWT can resolve frequencies more precisely and time resolutions more coarsely at lower frequencies. Fig. 3 shows 2D-DWT output AC, HC, VC and DC. The approximation coefficient is same as the input image, Horizontal coefficient sub band is giving the horizontal information of texture features, vertical coefficient sub band gives the vertical information of texture features similarly the diagonal sub band gives the diagonal texture features. Only HC, VC and DC are texture information is sufficient to analyze the texture features of inner trabecular bones micro-architecture of femur bone.

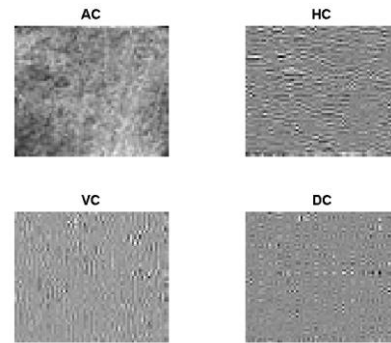


Fig. 3. First level 2D-DWT decomposition structure.

2D-DWT followed by Maximum Response Filter (MRF) based sub band image texture analysis is a technique used in order to extract meaningful texture features from an image as shown in Fig. 4, Multi-resolution analysis. In 2D-DWT, the image is decomposed into multiple frequency bands, representing different levels of detail or textures. This multi-resolution property is well-suited for texture analysis as textures often exhibit varying degrees of complexity at different scales. By applying the MRF at two sets of scale as [0.3, 0.6, 0.9] and [5, 10, 15] and filter size is technique after the 2D-DWT of HC, VC and DC, the most significant texture information from different sub bands can be extracted. MRF highlights regions with the maximum texture response, enhancing the representation of dominant texture patterns.

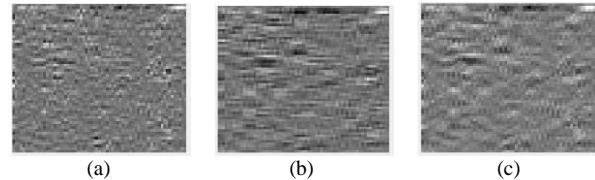


Fig. 4. First image texture (a) sum of MRFs scale at [0.3, 0.6, 0.9] (b) sum of MRFs scale at [5, 10, 15] (c) sum image of (a) and (b).

The matching is taking place between the supervised image database and test images based on SSD and ZNCC attributes and then decision is made on these attributes to decide whether the test ROI is abnormal (osteoporosis or osteopenia) or normal bone based on maximum ZNCC and minimum SDD.

According to the Table IV, different data cross-folding methods result in different confusion matrices. By applying cross-validation techniques, whether the model will perform well on unseen data and avoid over fitting or under fitting issues. As a result of supervised dataset for four folding, that is, 80 samples and the test dataset is 22 samples, the method achieves better results for four folding because of a greater number of supervised datasets. The system will have a larger number of texture patterns to decide whether an image is normal or abnormal, however osteoporosis and osteopenia bones both are included as abnormal in this proposed work. Actual positive sample (abnormal) set in test data set is 12 (osteoporosis and osteopenia) and actual negative sample set is 10 (normal). In the case of four folding, the model gives TP=12 and FP=0, which gives better results, however in normal case gives TN=9 out of 10 and FN=1.



TABLE IV. CONFUSION MATRIX FOR DIFFERENT FOLDING

Data Folding	TP	FP	TN	FN
Two-fold	11	3	4	4
Three-fold	12	2	5	3
Four-fold	12	0	9	1

Using graphical representation of non-normalized confusion matrix, system performance can easily assess at a momentary look as shown in Fig. 5 for four folding data. The diagonal of the confusion matrix (from top-left to bottom-right) represents correct predictions, while off-diagonal elements indicate misclassifications. This graphical representation helps in understanding the model strengths and weaknesses in terms of classifying different instances. The observation of four folding data from Table IV gives the TP=4+8=12 i.e., abnormal bone = Osteopenia (class-1) + Osteoporosis (class-2) and TN=10 i.e., normal bone (class-3). The system predicts better results with false negative was the only one sample, which is actually normal (healthy) bone showing it as abnormal.

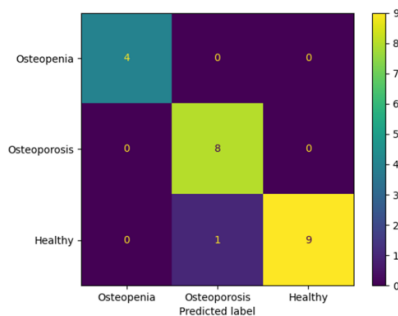


Fig. 5. Graphical representation confusion matrix for four folding data.

On all metrics measures, four folding achieves better results than the other folding, as revealed in Table V. Model performance metrics provide quantitative measures to evaluate how well a machine learning model is performing on a dataset. Metrics like accuracy, precision, recall, and other characteristics are essential for evaluating model. The choice of data folding (i.e., cross-validation method) can impact the way these metrics are computed.

TABLE V. MODEL PERFORMANCE METRICS FOR DIFFERENT DATA FOLDING

Data Folding	Precision	Recall	Specificity	F1 Score	Accuracy
Two-fold	78.57%	73.33%	57.14%	75.85%	68.18 %
Three-fold	85.71%	80.00%	71.42%	82.75%	77.27%
Four-fold	100%	93.33%	100%	96.54%	95.45%

Testing of the proposed model includes a general classifier technique in machine learning [31]: K-Nearest Neighbor (KNN), Discriminant Analysis (DA), Naive Bayes (NB), Decision Tree (DT), Support Vector Machine (SVM), and Random Forest (RF). For each classifier and proposed method, total twenty-one experiments were conducted including two, three, and four data folding. Comparing the system accuracy

for different classifiers is an essential step in selecting the most suitable classifier for a specific task. The accuracy metric provides an overall measure of how well the classifier performs in terms of correctly classifying instances. The Table VI shows that the proposed model is more accurate than other classification techniques, due to the effectiveness of the texture analysis method, so the system could be able to distinguish between normal and abnormal.

TABLE VI. FOUR-FOLD OF DATA COMPARISON OF SYSTEM ACCURACY FOR DIFFERENT CLASSIFIERS

Classifiers	Accuracy		
	Two-fold	Three-fold	Four-fold
KNN	58%	59%	61 %
DA	62%	64%	73 %
NB	60%	64%	66%
DT	67%	67%	68%
SVM	74%	75%	75%
RF	68%	68%	68%
<b>Proposed method</b>	<b>68.18 %</b>	<b>77.27%</b>	<b>95.45%</b>

The Table VII indicates that the proposed method for texture analysis performed better than other classification techniques, allowing the system to distinguish normal from abnormal behavior based on the performance evolution results.

TABLE VII. FOUR-FOLD COMPARISON OF PERFORMANCE EVALUATIONS FOR DIFFERENT CLASSIFIERS

Classifiers	Precision	Recall	Specificity	F1 Score	AU-ROC value
KNN	60 %	72 %	57%	65%	0.6128
DA	65%	72 %	60 %	68%	0.6789
NB	54%	60 %	67%	66%	0.6534
DT	60 %	60 %	60 %	60 %	0.5934
SVM	94 %	91%	93%	94%	0.9489
RF	92 %	92%	93 %	94 %	0.9393
<b>Proposed method</b>	<b>100%</b>	<b>93.33%</b>	<b>95.45%</b>	<b>96.54%</b>	<b>0.9659</b>

The An AU-ROC (Area Under the Receiver Operating Characteristic) plot is a graphical representation used to compare the performance of different classifiers in classification tasks. When comparing different classifiers using AU-ROC plots, the classifier with the highest AU-ROC value is generally considered as best performance. AU-ROC plot of the proposed system with two classification techniques is revealed in Fig. 6, in which the proposed occupies a larger area under the curve than the other two.

It is critical to compare the performance of different methods for determining osteoporosis in the femur bone when analyzing image data. Various methods or algorithms are employed to solve a particular problem, and comparing their performance helps in selecting the most effective one. As an illustration of the system performance for the different methods is as revealed in Table VIII, nevertheless proposed method performs better when data is folded four times.

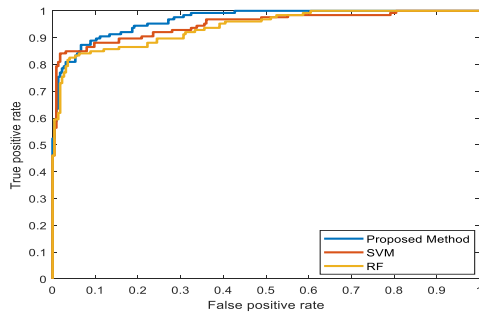


Fig. 6. An AU-ROC plots.

TABLE VIII. PERFORMANCE COMPARISON WITH DIFFERENT METHODS

Ref. No.	Specificity	F1 Score	Accuracy	Sensitivity	AU-ROC
[4]	88.24%	89.43%	88.50%	88.19%	89.01%
[15]	93.70%	93.39%	93.39%	93.90%	92.37%
[17]	84.20%	85.39%	91.39%	92.80%	86.50%
[22]	94.19%	90.28%	91.27%	80.29%	93.43%
Proposed Method	95.45%	96.54%	95.45%	93.33%	96.59%

## VI. CONCLUSION

The spectral based analysis of texture features of ROI femur X-ray images produces very good results because the 2D-DWT gives pixel intensity variation at different scale. Therefore, the texture features are helpful in obtaining the significant texture using MRFs. The various scales in image texture benefit from the MRFs, contributing to its effectiveness and the capability to process fine and coarse information efficiently. By calculating the two attributes as ZNCC and SSD of the test and supervised images in the matching process helps to classify the test image. The best matching between the test and the supervised database is considered as the maximum correlation and minimum sum difference which is exactly the test predicted classes. The proposed work achieves better osteoporosis detection with less error.

In future the evaluation can be made in spatial domain so that the system can be in contrast with proposed one however requiring extra database to boost the system performance. The femur X-ray images must be derived from a customary public dataset to make this work useful in medical disciplines for early detection of osteoporosis. A generalized X-ray image dataset for the femur is not available to make the proposed work more significant for detecting osteoporosis.

## REFERENCES

[1] LeBoff, M. S., S. L. Greenspan, K. L. Insogna, E. M. Lewiecki, K. G. Saag, A. J. Singer, and E. S. Siris. "The clinician's guide to prevention and treatment of osteoporosis." *Osteoporosis international* 33, no. 10 (2022): 2049-2102.

[2] Shen, Y., X. Huang, and J. Wu. "The global burden of osteoporosis, low bone mass, and its related fracture in 204 countries and territories, 1990–2019." *Front Endocrinol (Lausanne)* 13: 882241. (2022).

[3] Tarantino, Umberto, Giovanni Iolascon, Luisella Cianferotti, Laura Masi, Gemma Marcucci, Francesca Giusti, Francesca Marini et al. "Clinical guidelines for the prevention and treatment of osteoporosis:

summary statements and recommendations from the Italian Society for Orthopaedics and Traumatology." *Journal of orthopaedics and traumatology* 18, no. 1 (2017): 3-36.

[4] Yamamoto, N., S. Sukegawa, A. Kitamura, R. Goto, T. Noda, K. Nakano, K. Takabatake et al. "Deep learning for osteoporosis classification using hip radiographs and patient clinical covariates." *Biomolecules*. 2020; 10 (11)." 1534.

[5] Whitmarsh, Tristan, Ludovic Humbert, Mathieu De Craene, Luis M. Del Rio Barquero, and Alejandro F. Frangi. "Reconstructing the 3D shape and bone mineral density distribution of the proximal femur from dual-energy X-ray absorptiometry." *IEEE transactions on medical imaging* 30, no. 12 (2011): 2101-2114.

[6] Vijay, A., N. Shankar, C. Aroba Sahaya Liges, and M. Anburajan. "Evaluation of osteoporosis using CT image of proximal femur compared with dual energy X-ray absorptiometry (DXA) as the standard." In 2011 3rd International Conference on Electronics Computer Technology, vol. 3, pp. 334-338. IEEE, 2011.

[7] Pramudito, J. T., S. Soegijoko, T. R. Mengko, F. I. Muchtadi, and R. G. Wachjudi. "Trabecular pattern analysis of proximal femur radiographs for osteoporosis detection." *Journal of Biomedical & Pharmaceutical Engineering* 1, no. 1 (2007): 45-51.

[8] Supaporn, Kiattisin, and Chamnongthai Kosin. "Femur bone volumetric estimation from a single X-ray image for osteoporosis diagnosis." In 2006 International Symposium on Communications and Information Technologies, pp. 1149-1152. IEEE, 2006.

[9] Shankar, N., V. Sathagirivasan, A. Vijay, K. Kirthika, and M. Anburajan. "Evaluation of osteoporosis using radiographic hip geometry, compared with dual energy X-ray absorptiometry (DXA) as the standard." In 2010 International Conference on Systems in Medicine and Biology, pp. 259-264. IEEE, 2010.

[10] Sathagirivasan V., and M. Anburajan. "Analysis of trabecular proximal femur bone in diagnosing osteoporosis using digital x-ray: A comparison with DXA." In 2011 International Conference on Recent Trends in Information Technology (ICRITIT), pp. 869-574. IEEE, 2011.

[11] Mengko, Tati Rajab, and J. Tjandra Pramudito. "Implementation of Gabor filter to texture analysis of radiographs in the assessment of osteoporosis." In Asia-Pacific Conference on Circuits and Systems, vol. 2, pp. 251-254. IEEE, 2002.

[12] Reshmalakshmi, Chandrasekharan, and M. Sasikumar. "Fuzzy inference system for osteoporosis detection." In 2016 IEEE Global Humanitarian Technology Conference (GHTC), pp. 675-681. IEEE, 2016.

[13] Bobby, T. Christy. "Estimation of femur morphometric features for CBIR application." In 2017 Third International Conference on Biosignals, Images and Instrumentation (ICBSII), pp. 1-5. IEEE, 2017.

[14] Sahiti Lahari, M., and Anburajan M. Vijay. "Finite Element Analysis of Femur in the Evaluation of Osteoporosis." *Electronics Computer Technology (ICECT)*. In 2011 3rd International Conference on. IEEE Conference Proceedings, vol. 3, pp. 415-419. 2011.

[15] Shankar, N., S. Sathish Babu, and C. Viswanathan. "Femur bone volumetric estimation for osteoporosis classification using optimization-based deep belief network in X-ray images." *The Computer Journal* 62, no. 11 (2019): 1656-1670.

[16] Raghavendra Chinchansoor , Dr. Subhangi DC. "Machine Learning Technique for Osteoporosis Caused Bone Fracture Detection in Femur Bones from 2D X-Ray Images". © 2018 JETIR December 2018, Volume 5, Issue 12

[17] Nguyen, B.N.T.; Hoshino, H.; Togawa, D.; Matsuyama, Y. Cortical thickness index of the proximal femur: A radiographic parameter for preliminary assessment of bone mineral density and osteoporosis status in the age 50 years and over population. *CiOs Clin. Orthop. Surg.* 2018, 10, 149–156. [CrossRef] [PubMed]

[18] C. C. Yang, M. B. Nagarajan, M. B. Huber, J. C. Gamio, J. S. Bauer et al., "Improving bone strength prediction in human proximal femur specimens through geometrical characterization of trabecular bone microarchitecture and support vector regression," *Journal of Electronic Imaging*, vol. 23, no. 1, pp. 13013, 2014.

[19] Adams, M.; Chen, W.; Holdorf, D.; McCusker, M.W.; Howe, P.D.L.; Gaillard, F. Computer vs. human: Deep learning versus perceptual

- training for the detection of neck of femur fractures. *J. Med. Imaging Radiat. Oncol.* 2019, 63, 27–32. [CrossRef]
- [20] G. Chang, A. H. Cho, H. Rusinek, S. Honig, A. Mikheev et al., "Measurement reproducibility of magnetic resonance imaging-based finite element analysis of proximal femur microarchitecture for in vivo assessment of bone strength," *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 28, no. 4, pp. 407–412, 2015.
- [21] Yoon, Byung-Ho, Jang-Won Park, Chan Woo Lee, and Young Do Koh. "Different Pattern of T-Score Discordance between Patients with Atypical Femoral Fracture and Femur Neck Fracture." *Journal of Bone Metabolism* 30, no. 1 (2023): 87.
- [22] Hsieh, C. I., K. Zheng, C. Lin, L. Mei, L. Lu, W. Li, F. P. Chen et al. "Automated bone mineral density prediction and fracture risk assessment using plain radiographs via deep learning. *Nat Commun* 12: 5472." (2021).
- [23] Awal, Rabina, and Tanvir R. Faisal. "Multiple regression analysis of hip fracture risk assessment via finite element analysis." *Journal of Engineering and Science in Medical Diagnostics and Therapy* 4, no. 1 (2021): 011006.
- [24] Goel, Akash. "Discrete wavelet transform (DWT) with two channel filter bank and decoding in image texture analysis." *Int. J. Sci. Res* 3, no. 4 (2014): 391-397.
- [25] Fortuna-Cervantes, Juan Manuel, Marco Tulio Ramírez-Torres, Marcela Mejía-Carlos, José Salomé Murguía, José Martínez-Carranza, Carlos Soubervielle-Montalvo, and César Arturo Guerra-García. "Texture and Materials Image Classification Based on Wavelet Pooling Layer in CNN." *Applied Sciences* 12, no. 7 (2022): 3592.
- [26] Litimco, C.E.O., Villanueva, M.G.A., Yecla, N.G., Soriano, M.N. and Naval, P.C., 2013, March. Coral Identification Information System. In *IEEE International Underwater Technology Symposium (UT)* (pp. 1-6). IEEE 2013.
- [27] Soni, Pramod Kumar, Navin Rajpal, and Rajesh Mehta. "Road centerline extraction from VHR images using SVM and multi-scale maximum response filter." *Journal of the Indian Society of Remote Sensing* 49 (2021): 1519-1532.
- [28] Angeles, Maria Del Pilar, and Carlos G. Ortiz-Monreal. An attribute-based classification by threshold to enhance the data matching process. *Journal of applied research and technology* 17, no. 4 (2019): 272-284. .
- [29] quality sea-ice surface reconstruction from aerial images." *Remote Sensing* 11, no. 9 (2019): 1055. Vujović, Ž. "Classification model evaluation metrics." *International Journal of Advanced Computer Science and Applications* 12, no. 6 (2021): 599-606.
- [30] Sarker, Iqbal H. "Machine learning: Algorithms, real-world applications and research directions." *SN computer science* 2, no. 3 (2021): 160.
- [31] Sankara Subbu, Ramesh. "Brief Study of Classification Algorithms in Machine Learning." (2017).

# A Systematic Review on Blockchain Scalability

Asmaa Aldoubaee, Noor Hafizah Hassan, Fiza Abdul Rahim

Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia

**Abstract**—Blockchain is an exciting new technology that has garnered attention across multiple industries. This new technology offers several advantages, including decentralization, transparency, and immutability. However, several issues limit the effectiveness of this technology, such as scalability, interoperability, and privacy. A systematic review of blockchain scalability research was conducted using three primary databases: ACM, Science Direct, and IEEE. The review examined the state of the art in blockchain scalability, identifying the most important research trends and challenges. The solutions that have been established can be categorized into two main groups: those that pertain to block storage and those that pertain to the underlying blockchain mechanism. Numerous solutions were suggested for each main group. The most common proposed solutions for improving the scalability of blockchain networks in the literature are improving the consensus algorithm and using sharding. Most of the solutions were proof of concept and need more investigation in the future.

**Keywords**—Blockchain; scalability; sharding; consensus algorithm

## I. INTRODUCTION

Blockchain is another form of digital value exchange that has gained attention from different sectors [1]. The idea of blockchain was first introduced by Haber and Stornetta in 1991. Nakamoto later used blockchain in 2008 as the most well-known example: cryptocurrencies [2],[3]. This technology has found its way into various industries and applications, including banking, insurance, supply chain management, healthcare, identity verification, stock market analysis, IoT, energy, and intellectual property management. According to [4], the reasons for its popularity are its advantages, which include decentralization, security, immutability, efficiency, and transparency.

Despite the many advantages of blockchain technology, researchers and developers have identified several challenges and bottlenecks that need to be addressed before blockchain can be widely adopted, as [4] mentioned. Scalability is a significant challenge preventing the system from growing up. Scalability issue occurs for many reasons. According to [5], the main two are the blockchain mechanism and the block size.

To address this challenge, researchers and developers have made significant efforts to improve the scalability of blockchain technology [5]. Efforts can be categorized based on the areas they focus on. Some solutions aim to enhance the chain mechanism, while others, such as [6], concentrate on managing stored data.

This paper aims to review current blockchain scalability solutions and research trends systematically. This systematic literature review (SLR) briefly overviews blockchain

technology and its scalability problems. The different ways blockchain scalability has been addressed and the results of performance evaluations of these solutions are categorized. The paper identifies several potential areas for future research on blockchain scalability, such as improving the consensus mechanism, using sharding, and off-chain scaling.

This systematic review article consists of seven sections. Section I specifically covers the introduction and significance of this review. Section II provides a background about the blockchain and the scalability challenge. This is followed by Section III, highlighting the related reviews and surveys conducted on this topic. Section IV underlines the reasons for conducting this review. The following is Section V, which highlights the methodology of this review, including the research questions, study selection, and the inclusion and exclusion criteria. The results of this study, including the answers to the research questions, are discussed in Section VI. Lastly, this article is concluded in Section VII.

## II. BACKGROUND

### A. Blockchain

A blockchain is a chain of blocks that serves as a public ledger and contains complete records of all transactions committed [7], as illustrated in Fig. 1. The list of transactions in this chain expands as new blocks are added. This record of the list continues to grow.

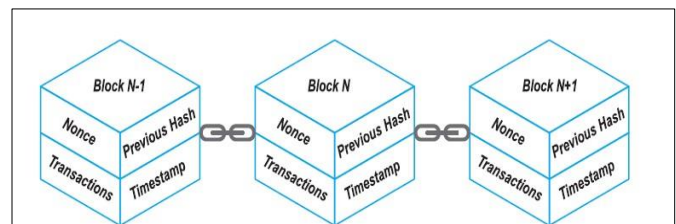


Fig. 1. A Standard structure of a blockchain [8].

Blockchain technology was first used in the field of cryptocurrencies in 2008. The success of blockchain technology in the world of cryptocurrencies has led other industries to explore and adopt it. The idea of blockchain was first introduced by Haber and Stornetta in 1991. Nakamoto later used blockchain in 2008 as the most well-known example, cryptocurrencies. This technology has found its way into various industries [12].

The success of blockchain technology is limited due to its inability to be implemented on a large scale [9]. In simple terms, scalability issues exist due to the limited block size and the current blockchain mechanism. This issue grows as the number of transactions increases, demanding additional nodes to maintain the network while also increasing the number of

steps required for the transaction to travel and attain full consensus with every node. For instance, according to [10], Bitcoin, which utilizes the Proof of Work (PoW) consensus algorithm, has a peak limit of processing only seven transactions per second.

Blockchain systems have lower throughput and latency performance than non-blockchain system [11]. The number of transactions completed per second is called throughput [10], while the delay between making a blockchain data request and getting a response to that request is referred to as latency.

This scalability issue has been studied, and many solutions have been proposed to enhance the ability to scale up the blockchain, which will be tackled later. It is not easier to propose a scalability solution because the features of blockchain will be affected, like decentralization and security [12]. Therefore, creating a trade-off between the proposed solution and the other related aspects is necessary, as illustrated in Fig. 2. The proposed solution must provide a trade-off between scalability, decentralization, and security. These characteristics can be challenging to balance, but it is essential to consider them when making decisions.

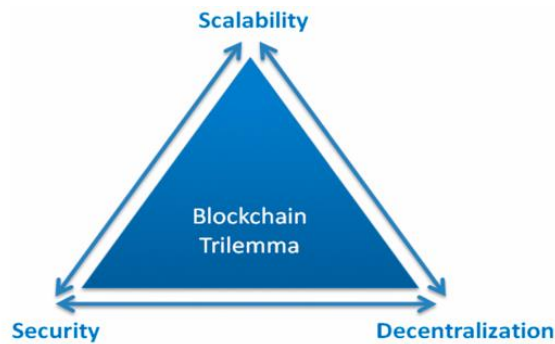


Fig. 2. Scalability trilemma [4].

### III. RELATED WORKS

With the growth of blockchain in various sectors, warning bells ring about the scalability issue. As a progression of this issue, researchers have thoroughly probed into this problem, and tens of papers have been published regarding this issue. Furthermore, scalability issues have been widely investigated recently. Tens of papers were published on this bottleneck. Moreover, reviews and surveys were conducted on this issue. Table I illustrates the relevant reviews and surveys on this SLR. Some related works focused on a special domain like healthcare [13] or the Internet of Medical Things [14]. This paper thoroughly explores blockchain scalability issues in general and evaluates proposed solutions to determine their effectiveness in practice rather than just as a proof of concept.

### IV. REASON FOR CONDUCTING SYSTEMATIC REVIEW

The reason for doing this Systematic Review is the critical importance of the scalability challenge within the blockchain. Extensive research has been conducted, and numerous attempts have been made to address the issue. However, there is still a need for enhancing and discovering more efficient methods to enhance the scalability of the blockchain. Therefore, within this

systematic review, we seek to highlight findings, identify gaps, and pave the way to discovering innovative, more effective methods to improve blockchain scalability. The output of this research may help to assist blockchain technology in evolving and changing, making it even more reliable, scalable, and flexible to meet the many demands of modern applications and industries.

### V. REVIEW METHOD

A systematic literature review identified, evaluated, and interpreted all available research related to a specific research question, topic area, or phenomenon of interest. The authors used guidance from [16], as a step and guide for doing the review, which served as a framework for their methodology. The goal of following this guidance was to ensure the review process was methodical and precise. The authors followed the suggested procedures, methods, and instructions in Fig. 3. An explanation for that will be provided.

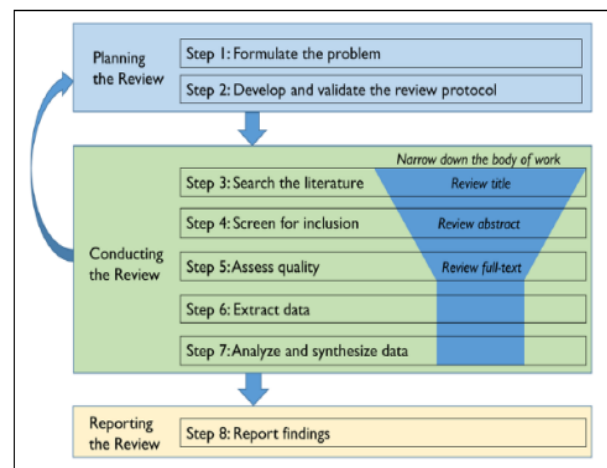


Fig. 3. Systematic review steps.

#### A. Research Question

The goal of this study is to investigate the scalability challenge of blockchain. The following research questions were formulated to conduct the investigation:

- RQ1: Where are the studies on the scalability of blockchain being conducted?

This research question aims to understand the current research and study trends on blockchain scalability.

- RQ2: How the scalability issues in blockchain have been addressed? What are the techniques and aspects used to address the scalability?

The second research question is to identify the current approaches and solutions by looking at the techniques and aspects used to solve blockchain scalability.

- RQ3: How did the proposed solutions succeed in achieving scalability?

The last research question is to evaluate the impact of prior efforts and pave the way for more effective and scalable blockchain systems in the future.

TABLE I. RELATED REVIEWS AND SURVEYS

Ref	Year	Title	Summary	Main Future direction
[12]	2021	A systematic review of blockchain scalability: Issues, solutions, analysis, and future research	In this review, the available solutions were categorized according to their performance in three areas: writing, reading, and storage.	This paper proposes integrating two or more scalability solutions to create a more effective and secure scaling solution. The goal is to enhance the read performance of the blockchain and optimize the query language. Additionally, it recommends implementing more robust cross-shard communication methods.
[14]	2021	A Survey on Blockchain-Based IoMT Systems: Towards Scalability	The survey presented the various factors that can affect a blockchain's ability to scale, whether directly or indirectly. Additionally, it categorized the potential solutions into two types: on-chain and off-chain.	<ul style="list-style-type: none"> <li>• They recommend two methods to solve the scalability issue: either through on-chain or off-chain solutions.</li> </ul>
[15]	2020	A Review on Scalability of Blockchain	<ul style="list-style-type: none"> <li>• The review outlined the key approaches and technologies put out to address the scalability issue in the blockchain.</li> <li>• There are three primary factors that contribute to blockchain scalability bottlenecks. These include performance inefficiency, significant confirmation delays, and function extension limitations.</li> </ul>	<p>The review provided suggestions for further studies as below:</p> <ul style="list-style-type: none"> <li>• Studying a large-scale, high-performance peer-to-peer (P2P) network. Without advancements in network technology, enhancing the performance of blockchain systems will remain challenging.</li> <li>• A high-performance programmable computing engine is crucial for utilizing various smart contracts that are written in different programming languages.</li> </ul>
[13]	2020	Scalability Challenges in Healthcare Blockchain System—A Systematic Review	<ul style="list-style-type: none"> <li>• Defined the main reasons leading to healthcare scalability issues: the block size, huge amounts of data, the number of nodes, and the consensus protocol.</li> <li>• It provided a map of the main 16 proposed solutions according to the reasons. This review covers 16 solutions that fall into two main categories: storage optimization and blockchain redesign. There are three solutions for storage optimization and 13 solutions for blockchain redesign, including blockchain modeling, read and write mechanisms and bi-directional network.</li> </ul>	NA
[16]	2022	Scalable blockchains — A systematic review	<ul style="list-style-type: none"> <li>• This review classified the current solutions into solutions related to payment Channel Networks like lightning networks, sharding, blockchain delivery networks, hardware-assisted networks, Parallel Processing, and blockchain redesigning.</li> <li>• It highlighted the sharding as the main potential solution.</li> </ul>	<p>This review offers these recommendations.:</p> <ul style="list-style-type: none"> <li>• Developing new consensus algorithms.</li> <li>• Exploring the use of off-chain solutions.</li> <li>• Investigating the potential of sharding and sidechains to improve the scalability of blockchains.</li> <li>• Developing new metrics evaluate scalability of blockchains.</li> <li>• Investigating the impact of blockchain scalability on various application domains.</li> <li>• Developing new tools and frameworks to facilitate the development and deployment of scalable blockchain applications.</li> </ul>
[4]	2021	Systematic Literature Review of Challenges in Blockchain Scalability.	<ul style="list-style-type: none"> <li>• This review analyzed the main factors that affected the scalability of blockchain: the number of transactions per second, and the consensus mechanism and how it affects the scalability.</li> <li>• The review explores the on-chain and off-chain solutions to the scalability issue. Consensus algorithm and sharding were the most important solutions.</li> </ul>	NA
[6]	2020	Solutions to Scalability of Blockchain: A Survey	The scalability problem with blockchain systems is discussed in this study along with a number of suggested solutions. Some of the suggested solutions include sharding, sidechains, cross-chain solutions, DAG-based solutions, and off-chain solutions including payment channels and state channels. These solutions are grouped into many typical blockchain layers. The authors also go through alternative consensus techniques and how they can help blockchain systems become more scalable, including Proof of Work (PoW), Proof of Stake (PoS), and Delegated Proof of Stake (DPoS).	Future work will involve building more effective and safer sharding approaches, researching the possibilities of off-chain solutions like payment channels and state channels, and examining alternative consensus mechanisms that might increase the scalability of blockchain systems. The authors also advise investigating the use of artificial intelligence and machine learning methods to enhance the functionality of blockchain systems. Finally, they advise looking into the possibility of combining blockchain with other cutting-edge technologies like edge computing and the Internet of Things (IoT).

### B. Study Selection

Through a systematic review process, various research works have been found in published papers to explore the research trends and state-of-the-art advancements in blockchain scalability. The initial search used ACM, Science Direct, and IEEE databases. They were chosen based on the availability of these papers as full texts through the Universiti Teknologi Malaysia library. This phase was conducted by using the following search string: (“blockchain scalability” OR “scalable blockchain”), which is summarized in Table II. After conducting an initial search of the three databases, a total of 146 papers were found. These papers went through different stages of filtering, based on the criteria shown in Fig. 2. Finally, the 35 studies were reviewed to identify trends and advancements in blockchain scalability.

### C. Inclusion and Exclusion Criteria

Inclusion criteria are established in the SLR to find papers related to the research goals. These standards guarantee that studies directly related to the study issue are included, whereas studies that do not conform to the established standards or are irrelevant to the research emphasis are excluded. The paper selection process is illustrated in Fig. 4. The main inclusion criteria for this study are:

- Being published in English.

- Being cited at least once.
- Describing the scalability issue on a blockchain.
- Providing solutions for scalability obstacles.

The exclusion criteria are:

- Papers that are written in another language rather than English
- Papers in which the concept is not described clearly.
- Do not specifically address scalability challenges in the blockchain.
- Papers that are not cited in other studies.

TABLE II. PAPERS SELECTION FROM THE DATABASES

Data Source	Documentation	
	No. of Articles Found on Primary Search	Selected Articles
ACM (Journal)	304	9
Science Direct	307	6
IEEE	109	20
Total	720	35
Total		35

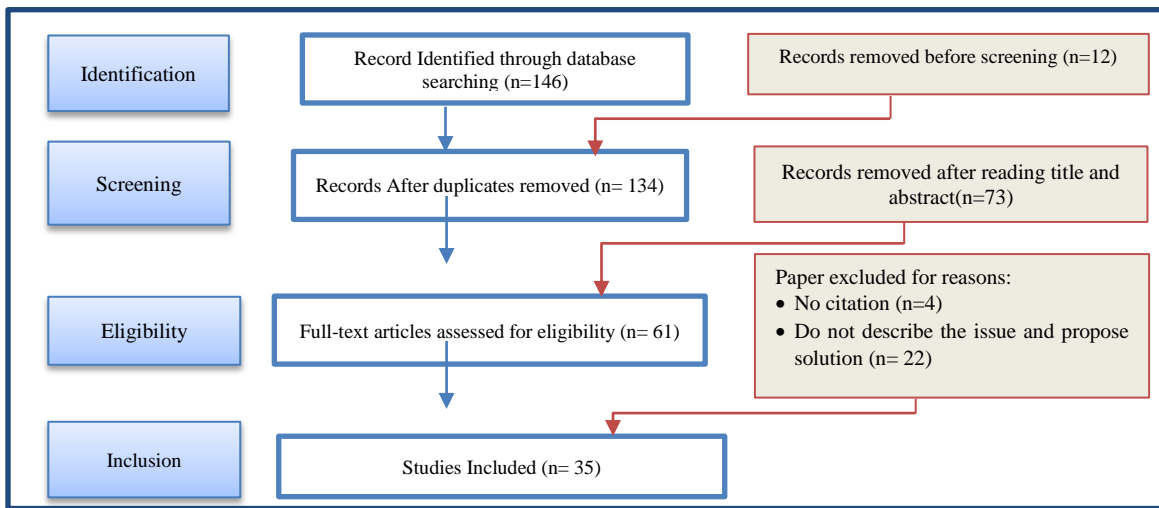


Fig. 4. Papers selection process.

## VI. RESEARCH FINDINGS

In this section, we will discuss the results of our systematic review to answer the research questions listed below.

RQ1: Where are the studies on the scalability of blockchain being conducted?

The issue of scalability has captured the attention of researchers, resulting in the publication of numerous papers in both journals and conferences. These studies explained the reasons that lead to the scalability issue, and accordingly, Numerous techniques have been proposed to address the scalability problem in blockchain from various perspectives.

Previous reviews on this issue have commonly classified the solutions based on their relationship to the blockchain, distinguishing between on-chain and off-chain approaches. On-chain solutions primarily focus on improving scalability by modifying elements within the blockchain, while off-chain solutions prefer to conduct transactions outside the network [4] or based on the purpose or objective of the proposed solution [11].

The approaches presented in this systematic review were broken down into two groups: related to the Blockchain mechanism and related to block storage. The largest share of proposed solutions dealt with the mechanism of the blockchain, followed by those that dealt with block storage, as illustrated in Fig. 5.

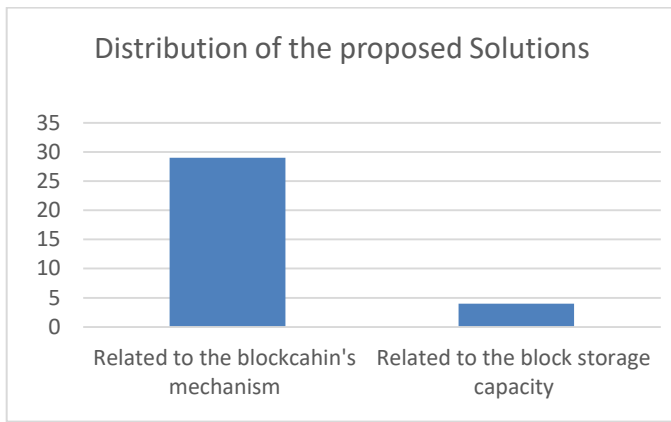


Fig. 5. Distribution of the founded solutions.

RQ2: How the scalability issues in blockchain have been addressed? What are the techniques and aspects used to address the scalability?

As mentioned previously, there are two main classifications in this systematic literature review. Under each main classification, several techniques were employed to address the issue of scalability. Fig. 6 illustrates the classification, whereas Fig. 7 illustrates the distribution of the proposed technique. The techniques found under each group are illustrated below:

A. Related to the Blockchain's Underlying Mechanism

1) *Optimizing the consensus algorithm*: The consensus protocol is a crucial component of blockchain that facilitates the creation of new blocks and the maintenance of the network. It involves reaching an agreement among network users on sustaining the network. The consensus protocol outlines the process for selecting the author of a new block [10].

Various applications have been implemented to utilize these techniques to improve the scalability of the blockchain. Based on the findings of this SLR, the consensus protocol is most frequently discussed to propose solutions to the scalability issue. Many studies concluded that the ineffectiveness of the consensus protocol primarily causes the main blockchains' scalability problems. So, to address the

scaling issue, researchers have looked for novel consensus methods [4].

The consensus algorithm has been developed from different aspects: reducing the complexity [17], scaling according to the incoming traffic rate [18], and using checkpoints that allow the block to have its own hash chains to add more transactions [19]. The implementation results encourage more investigation and development to generate a better scaling rate. Table III shows the major consensus solutions found in this review.

2) *Using sharding*: The concept of sharding has gained popularity, particularly because it has been successfully applied in the field of databases in the past. It is becoming more well-known as one of the viable ways to improve blockchain performance [16]. Sharding has been used widely as a solution for scalability bottlenecks in the blockchain. It enables blockchains to expand effectively. Furthermore, it divides the workload across different subsets of nodes to handle different parts of the blockchain to reduce the overhead of consensus protocol [20] [21]. As a result, each node is no longer required to process the whole network's transactional load. Each node only stores the information relevant to the partition or shard responsible for it. Among all the other scaling solutions, sharding seems the most effective solution as it holds the core functionalities of blockchain with it. Each shard works like a separate blockchain network, operating completely as Satoshi Nakamoto envisioned a blockchain to work[22]. However, applying sharding to the blockchain presents several difficulties, and there are no fully prepared methods to increase the blockchain's scalability. For instance, security inside the sub-shards and the inter-shard communication method are issues [23]. Furthermore, researchers found a problem with sharding with the node generation process [24]. Finally, many shard-based solutions were proposed with different benefits for increasing the scalability of the blockchain. Many are still theoretical concepts that need real implementation for further evaluation, as is clear in Table IV.

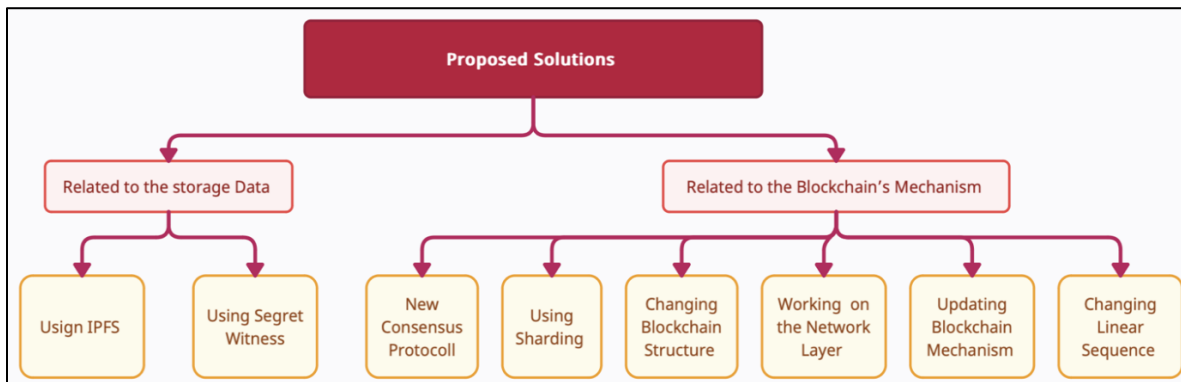


Fig. 6. Proposed solutions classifications.



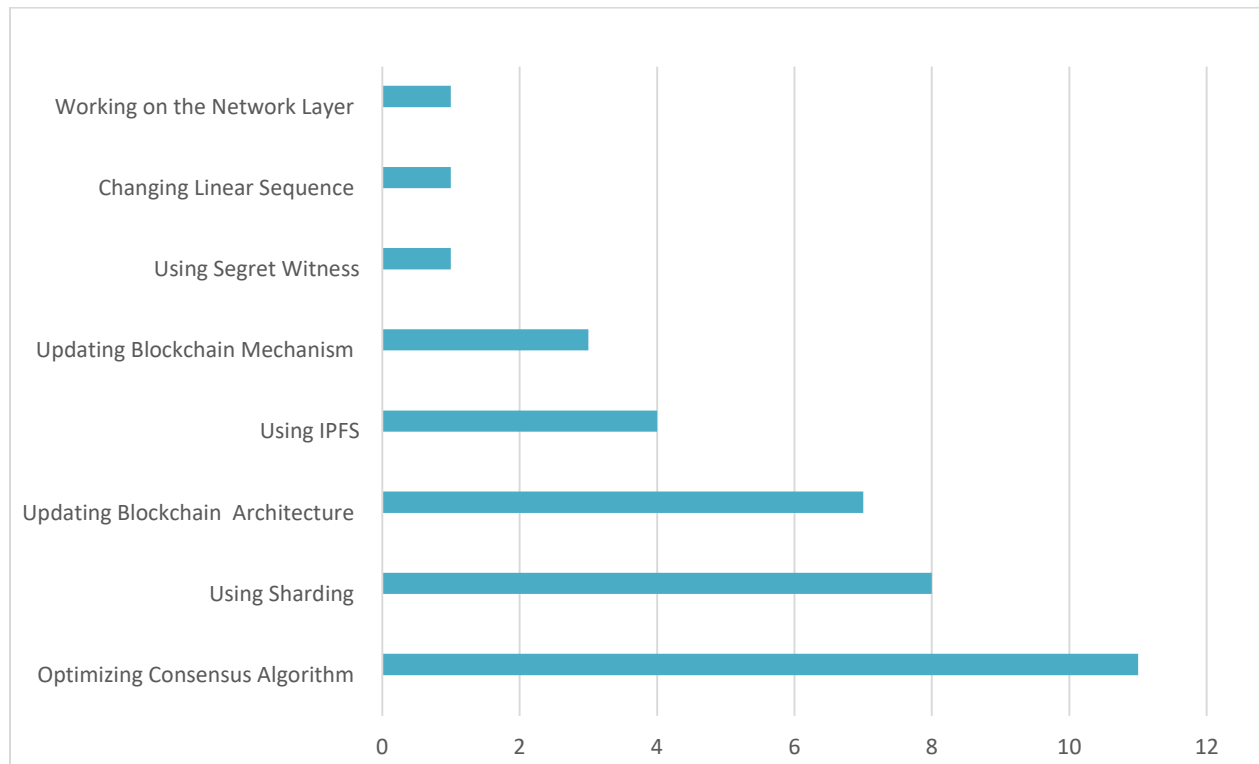


Fig. 7. The distribution of the proposed solution.

TABLE III. MAJOR SOLUTIONS BASED ON CONSENSUS ALGORITHM

Ref	Year	Protocol	Improve scalability by	Goal	Implementation	Results
[17]	2022	Byzantine Algorithm Group	Reducing the communication complexity	reducing complexity from $O(n^2)$ to $O(n_2^3)$	-	-
[25]	2019	Byzantine Fault Tolerant	Randomly select root	Selecting root randomly will reduce the mining time.	-	Providing constant latency
[18]	2021	Dynamic Proof of Work (PoW)	mining Scaling according to the IoT traffic	Checkpoint mechanism for mining that	NA	
[26]	2019	Byzantine Fault Tolerant	Reduce communication on each node	Cloud service integrated with the blockchain	-	-
[27]	2018	Delegate Proof of Stake (PoS)	Randomize delegated		-	-
[28]	2021	Hybrid protocol (Pow) and (PoS)	Transactions from external sources can be processed through this protocol, which then sends the final outcomes back to the main layer for storage in the distributed ledger.	The Hybrid Lightning Protocol can scale blockchain networks to handle more transactions.	√	maximized the throughput to 1,668,000 Transactions per second
[29]	2019	Distributed Time-Based consensus algorithm	Proposed algorithm that could minimize the processing overhead	To minimize the confirmation time	√	Successfully decreasing the processing time

TABLE IV. SUMMARY OF SHARDING-BASED SOLUTIONS

Ref.	Idea	Improve scalability by	Goal	Implementation	Results
[30]	Muti level sharding	Providing architecture for multi-level sharding and proposed a mechanism for interacting between sub-nodes	Enabling efficient cross-chain transactions in high scalability and extensibility.	NA	NA
[31]	Reducing the inter shard communication	Dividing the assignments, validation, verification, & storage responsibilities between nodes	Minimizing the need for inter-shard communication	NA	NA
[32]	Using the Verification Random Function	Utilizing a sharding strategy to share the random values created by participating nodes within a smaller group of nodes rather than directly among all node	Decreasing the computation and communication overhead	NA	NA
[20]	The scheme simultaneously achieves linear scaling in throughput, storage efficiency, & security.	Each node stores and computes in a coded shard of the same size that is generated by linearly mixing uncoded shards.	Achieving throughput efficiency as well as improving security.	Simulation	performing better than both uncoded Sharding and complete replication in throughput, storage optimization, and security.

3) *Changing the structure of blockchain:* The architecture of blockchain has a high computational complexity and needs a considerable amount of computing and storage space. These characteristics make it difficult to scale up. From that angle, many research studies proposed solutions to redesigning blockchain, like using a sub-chain [24]. Additionally, other researchers defined three types of blocks, and according to these types, block generation and consensus were updated [33]. However, this review observation shows that this solution has received relatively fewer citations. Therefore, we recommend conducting further studies soon to explore and investigate this area deeply.

4) *Improving the scalability from the second network layer:* Working on the network layer is an effective scaling option for blockchain. However, as it is an off-chain solution, it lacks the fundamental features of blockchain technology [22]. It became obvious how much scalability could be increased by using second-layer state channels [34]. The second layer channel works as an extra channel that can increase transactions per second and improve blockchain scalability by bypassing the consensus process.

5) *Updating the blockchain mechanism:* The suggested mechanism and solution focused on the activity inside the blockchain network, such as message passing between nodes. This reduced the amount of data held in each block and improved its scalability [35]. One way to achieve this is by altering the method of selecting the neighboring block [26]. The evaluation of these solutions shows better scalability in terms of reducing the time of confirmation.

6) *Changing the linear sequence of the blockchain:* Previous solutions have considered factors such as block size, block generation, and consensus. These solutions operate within the linear sequence of the blockchain network. Alternatively, altering the order of the block sequence may improve scalability [9]. Researchers suggested using a Directed Acyclic Graph-based blockchain model, which could enhance the scalability of large-scale networks [9]. A single solution was found in this review as a suggestion for enhancing the scalability. Furthermore, due to the limited

amount of published experimental findings and open-source implementations, this solution needs more investigation in the future.

### B. Related to Block Storage Capacity

One of the main factors affecting blockchain's scalability is the block storage capacity. This problem has been addressed with several proposals that try to expand the space available within blocks and permit more transactions. Here are some significant solutions that were discovered in this SLR:

1) *Using Segregated Witness (SEGwit):* Adding more space to the block allows more transactions to be done as SEGwit [36]. The concept involves restructuring transactions through forking, resulting in a four-fold increase in block size. Additionally, the block signature will be kept separate from the block, allowing for improved scalability and increased transaction capacity.

2) *Using Interplanetary File System (IPFS):* Many proposed solutions use InterPlanetary File System (IPFS) for storing data [37], while others use it for storing transactions [38]. One way to improve scalability is by using IPFS, which reduces the amount of data stored in a block by only including the hash value while the actual data is stored in IPFS. According to the analysis, this technology could play a significant role in scaling blockchain without affecting the core mechanism of blockchain [39].

RQ3: How did the proposed solutions succeed in achieving scalability?

This review found that the available studies have improved scalability by analyzing the factors that caused this obstacle. These factors were the block storage capacity and the mechanism of the blockchain. Most of the studies that suggested solutions could enhance scalability by dealing with these factors. Furthermore, some studies' initial results show incremental scalability improvement in throughput and latency. The studies contributed to enhancing scalability by analyzing the factors and providing theoretical-based solutions. The available solutions can serve as a roadmap for achieving improved scalability. Most of the studies still prove the

concept, lacking implementation in real scenarios. Scalability bottlenecks in blockchain still need further investigation. Table V summarizes all the research articles included in this SLR.

TABLE V. SUMMARY OF THE SELECTED STUDIES

Library	Ref.	Aspect	Simulation	Implementation	Improve scalability idea	Domain
ACM	[40]	Scalable consensus algorithm (BFT) and integration blockchain into cloud-based services	√	×	reducing the intra-plant communication complexity from $O(n^2)$ to $O(n)$ .	Industrial Plant
	[17]	The new consensus algorithm for reducing the complexity	√	×	reducing the communication complexity from $O(n^2)$ to $O(n^3)$	Commercial Blockchain
	[38]	Using IPFS	√	×	A theoretical manner for distributing transactions between on-chain and off-chain. Only the hash address is stored in the block, and using IPFS for storing the transactions.	Blockchain-base crypto computing
	[41]	New architecture	×	×	The maximum workload that may be handled varies with the number of nodes which increases the scalability by preventing broadcast in all cases	NA
	[9]	The DAG-based model includes greater scalability and lower transaction fees	×	×	Improving linear sequence and Separating the workload into a different level	NA
	[42]	New consensus algorithm : Proof of Property	×	×	Allowing participants to validate the transaction without downloading the complete blockchain which enhances the storage.	NA
	[21]	Using Sharding + enhancing consensus algorithm	√	√	The consensus protocol in each shard has been improved to achieve more than 3,000 transactions per second, resulting in increased effectiveness. This has been implemented across multiple shards.	NA
	[27]	Using randomize consensus algorithm for subchains	×	×	By using subchain technology, scalability is enhanced as it allows for additional blockchain nodes to become block producers and receive rewards.	NA
IEEE	[30]	Sharding	×	×	By facilitating efficient cross-chain transactions, multi-level sharding can be enabled to achieve high scalability and extensibility.	NA
	[43]	Using New architecture	√	×	Integrating with cloud storage for storing transactions to solve storage	NA
	[37]	Using IPFS	×	×	Improving scalability by utilizing IPFS to store patient records outside of the blockchain.	Medical Records
	[32]	Randomness protocol via sharding	×	×	Eliminating the use of heavy cryptography	Large-scale IoT applications
	[18]	Using new consensus algorithm	√	×	Using dynamic Proof Of Work consensus with checkpoint mechanism for mining Scaling according to the IoT traffic	Industrial Internet of Things
	[26]	Algorithm for choosing the block's neighbor	×	√	Reducing propagation time by 20-40%	NA
	[35]	using the data compression scheme	×	√	Reducing the data on the block by compression. Scalability improved by reducing the creation time .	Trading Platforms
[33]	New Blockchain Architecture	×	√	Reducing the consensus complexity will lead to better scalability.	NA	
[20]	Sharding with Lagrange-Coded Computing	×	×	We are enabling scaling without increasing the number of nodes by reducing the node's storage workload and increasing the verification outside the system.	NA	
[44]	Blockchain Architecture	×	√	They are reducing the resource	Massive devices of	

					utilization by updating the architecture for better scalability rate.	IoT
	[45]	Blockchain Architecture focuses on all layers	×	√	Reducing the storage need which operates via DHT	NA
	[25]	Consensus algorithm	×	√	Randomly selecting committees to improve on the quadratic message complexity	
	[46]	New Blockchain System	×	√	Reducing transaction storing by enabling SQL with Hadoop	Big Data
	[47]	Randomize the method for generation nodes by using master node technology	×	×	Raising the number of TP/S	Distributed Apps
	[24]	Scheme with the algorithm for node classification	√	×	Improving the way of nodes production	Information Blockchain
	[19]	Architecture for horizontal scalability	×	√	Delaying the transaction verification to increase the throughput by using checkpoint blocks which increase the number of created nodes.	NA
	[22]	Sharding	×	×	Enhancing the Sharding process by to be more effective and secure	NA
	[48]	Sharding with plasma	√	×	Increasing scalability by making the transactions process parallelly.	NA
	[34]	Second layer Network	×	×	Eliminating some of the transactions from the consensus process by passing their registration in the general ledger	NA
	[39]	IPFS	×	×	Decreasing the storing bloating issue by using IPFS	NA
	[31]	Sharding	×	√	Reducing Inter-shard communication	IoT
Science Direct	[28]	Hybrid consensus algorithm	×	×	Increasing scalability by doing validation and all transaction stored in the second layer and only the final transaction recorded in the first layer to raise throughput to 1,668,000 TP/S	NA
	[29]	Distributed Time Consensus algorithm	√	×	Reducing the mining time	IoT – smart home
	[49]	IPFS	×	√	Storing only the hash data in the block instead of the data itself	Medical Records
	[50]	Sharding + Microservice architecture	×	√	Increasing the throughput by decreasing the communication overhead	Big Data
	[51]	Architecture	√	×	Increasing throughput by splitting and payment to multiple channels.	Cryptocurrencies
	[52]	IPFS	×	×	Decreasing the propagation time	Electronic Health Records

## VII. CONCLUSION

In this paper, a systematic review was conducted to define the state of the art on blockchain scalability issues. The researchers attempted to analyze the studies through the literature extracted from the three databases. Introducing a solution for blockchain scalability bottleneck is not a simple process due to the complexity of the blockchain architecture and the distributed nature. Thus, proposed solutions must provide a trade-off between keeping the blockchain scaling and keeping its robust features like the decentralization and security. The consensus algorithm is the most critical component developed by researchers. Moreover, the Proposed solutions show improved scalability regarding throughput and latency. Additionally, sharding techniques come second as a critical solution that could significantly enhance scalability. Finally, most of the solutions are still proof of concept, and the recommendation is to apply these solutions in real scenarios for

the best investigation and analysis. Blockchain's scalability problem is a major obstacle to its widespread adoption. More research and better solutions must be developed to make widespread use of blockchain possible.

## ACKNOWLEDGMENT

The study was funded by the Encouragement Research Grant (Vote No. Q.K130000.3856.20J96) awarded by Universiti Teknologi Malaysia.

## REFERENCES

- [1] M. A. Hussain, M. S. Abd Latiff, S. H. H. Madni, R. Z. Raja Mohd Rasi, and M. F. I. Othman, "Concept of Blockchain Technology," *International Journal of Innovative Computing*, vol. 9, no. 2, pp. 51–57, 2019, doi: 10.11113/ijic.v9n2.238.
- [2] S. Haber and W. S. Stornetta, "How to Time-Stamp a Digital Document BT - Advances in Cryptology-CRYPTO' 90," pp. 437–455, 1991, [Online]. Available: [https://link.springer.com/content/pdf/10.1007/3-540-38424-3\\_32.pdf](https://link.springer.com/content/pdf/10.1007/3-540-38424-3_32.pdf)

- [3] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System," *Artif Life*, vol. 23, no. 4, pp. 552–557, 2017, doi: 10.1162/ARTL\_a\_00247.
- [4] D. Khan, L. T. Jung, and M. A. Hashmani, "Systematic Literature Review of Challenges in Blockchain Scalability," *Applied Sciences*, 2021.
- [5] B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering - A systematic literature review," *Inf Softw Technol*, vol. 51, no. 1, pp. 7–15, 2009, doi: 10.1016/j.infsof.2008.09.009.
- [6] Q. Zhou, H. Huang, Z. Zheng, and J. Bian, "Solutions to Scalability of Blockchain: a Survey," *IEEE Access*, vol. 8, pp. 16440–16455, 2020, doi: 10.1109/ACCESS.2020.2967218.
- [7] B. Badr, *Blockchain By Example*. Birmingham : Packt, 2018.
- [8] S. K. Dwivedi, P. Roy, C. Karda, S. Agrawal, and R. Amin, "Blockchain-Based Internet of Things and Industrial IoT: A Comprehensive Survey," *Security and Communication Networks*, vol. 2021, 2021, doi: 10.1155/2021/7142048.
- [9] Q. Wang, "Improving the scalability of blockchain through DAG," *Middleware 2019 - Proceedings of the 2019 20th International Middleware Conference Doctoral Symposium, Part of Middleware 2019*, pp. 34–35, 2019, doi: 10.1145/3366624.3368165.
- [10] A. I. Sanka and R. C. C. Cheung, "A systematic review of blockchain scalability: Issues, solutions, analysis and future research," *Journal of Network and Computer Applications*, vol. 195, no. December 2020, p. 103232, 2021, doi: 10.1016/j.jnca.2021.103232.
- [11] A. I. Sanka, M. H. Chowdhury, and R. C. C. Cheung, "Efficient High-Performance FPGA-Redis Hybrid NoSQL Caching System for Blockchain Scalability," *Comput Commun*, vol. 169, no. January, pp. 81–91, 2021, doi: 10.1016/j.comcom.2021.01.017.
- [12] A. I. Sanka and R. C. C. Cheung, "A systematic review of blockchain scalability: Issues, solutions, analysis and future research," *Journal of Network and Computer Applications*, vol. 195, no. September, p. 103232, 2021, doi: 10.1016/j.jnca.2021.103232.
- [13] A. A. Mazlan, S. M. Daud, S. M. Sam, H. Abas, S. Z. A. Rasid, and M. F. Yusof, "Scalability Challenges in Healthcare Blockchain System-A Systematic Review," *IEEE Access*, vol. 8, Institute of Electrical and Electronics Engineers Inc., pp. 23663–23673, 2020. doi: 10.1109/ACCESS.2020.2969230.
- [14] A. Advavoudi Jolfaei, S. F. Aghili, and D. Singelee, "A Survey on Blockchain-Based IoMT Systems: Towards Scalability," *IEEE Access*, vol. 9, pp. 148948–148975, 2021, doi: 10.1109/ACCESS.2021.3117662.
- [15] D. Yang, C. Long, H. Xu, and S. Peng, "A Review on Scalability of Blockchain," *ACM International Conference Proceeding Series*, pp. 1–6, 2020, doi: 10.1145/3390566.3391665.
- [16] M. H. Nasir, J. Arshad, M. M. Khan, M. Fatima, K. Salah, and R. Jayaraman, "Scalable blockchains — A systematic review," *Future Generation Computer Systems*, vol. 126, pp. 136–162, 2022, doi: 10.1016/j.future.2021.07.035.
- [17] J. Wu and N. Jiang, "SEGBFT: A Scalable Consensus Protocol for Consortium Blockchain," *ACM International Conference Proceeding Series*, pp. 15–21, 2022, doi: 10.1145/3532640.3532643.
- [18] U. Javaid and B. Sikdar, "A Checkpoint Enabled Scalable Blockchain Architecture for Industrial Internet of Things," *IEEE Trans Industr Inform*, vol. 17, no. 11, pp. 7679–7687, 2021, doi: 10.1109/TII.2020.3032607.
- [19] K. Cong and J. Pouwelse, "A Blockchain Consensus Protocol With Horizontal Scalability," *IEEE*, 2018.
- [20] S. Li, M. Yu, C. S. Yang, A. S. Avestimehr, S. Kannan, and P. Viswanath, "PolyShard: Coded Sharding Achieves Linearly Scaling Efficiency and Security Simultaneously," *IEEE International Symposium on Information Theory - Proceedings*, vol. 2020-June, pp. 203–208, 2020, doi: 10.1109/ISIT44484.2020.9174305.
- [21] H. Dang, T. T. A. Dinh, D. Lohin, E. C. Chang, Q. Lin, and B. C. Ooi, "Towards scaling blockchain systems via sharding," *Proceedings of the ACM SIGMOD International Conference on Management of Data*, pp. 123–140, 2019, doi: 10.1145/3299869.3319889.
- [22] A. Chauhan, O. P. Malviya, M. Verma, and T. S. Mor, "Blockchain and Scalability," *Proceedings - 2018 IEEE 18th International Conference on Software Quality, Reliability, and Security Companion, QRS-C 2018*, pp. 122–128, 2018, doi: 10.1109/QRS-C.2018.00034.
- [23] R. Zhijie and Peter. Zhou, "What does 'scalability' really mean in Blockchain?" Accessed: Dec. 01, 2022. [Online]. Available: <https://medium.com/vechain-foundation/what-does-scalability-really-mean-in-blockchain-b8b13b3181c6>
- [24] X. Hao, P. L. Yeoh, T. Wu, Y. Yu, Y. Li, and B. Vucetic, "Scalable Double Blockchain Architecture for IoT Information and Reputation Management," *7th IEEE World Forum on Internet of Things, WF-IoT 2021*, pp. 171–176, 2021, doi: 10.1109/WF-IoT51360.2021.9595791.
- [25] M. M. Jalalzai, C. Busch, and G. G. Richard, "Proteus: A scalable BFT consensus protocol for blockchains," *Proceedings - 2019 2nd IEEE International Conference on Blockchain, Blockchain 2019*, pp. 308–313, 2019, doi: 10.1109/Blockchain.2019.00048.
- [26] K. Wang and H. S. Kim, "FastChain: Scaling Blockchain System with Informed Neighbor Selection," in *2019 IEEE International Conference on Blockchain (Blockchain)*, 2019, pp. 376–383. doi: 10.1109/Blockchain.2019.00058.
- [27] X. Fan and Q. Chai, "Roll-DPos: A randomized delegated proof of stake scheme for scalable blockchain-based Internet of Things systems," *ACM International Conference Proceeding Series*, pp. 482–484, 2018, doi: 10.1145/3286978.3287023.
- [28] A. I. Fajri and F. Mahananto, "Hybrid lightning protocol: An approach for blockchain scalability issue," *Procedia Comput Sci*, vol. 197, pp. 437–444, 2021, doi: 10.1016/j.procs.2021.12.159.
- [29] A. Dorri, S. S. Kanhere, R. Jurdak, and P. Gauravaram, "LSB: A Lightweight Scalable Blockchain for IoT security and anonymity," *J Parallel Distrib Comput*, vol. 134, pp. 180–197, 2019, doi: 10.1016/j.jpdc.2019.08.005.
- [30] Y. Yu, R. Liang, and J. Xu, "A scalable and extensible blockchain architecture," *IEEE International Conference on Data Mining Workshops, ICDMW*, vol. 2018-Novem, pp. 161–163, 2019, doi: 10.1109/ICDMW.2018.00030.
- [31] S. R. Niya, R. Beckmann, and B. Stiller, "DLIT: A Scalable Distributed Ledger for IoT Data," *2020 2nd International Conference on Blockchain Computing and Applications, BCCA 2020*, pp. 100–107, 2020, doi: 10.1109/BCCA50787.2020.9274456.
- [32] G. Wang and M. Nixon, "RandChain: Practical Scalable Decentralized Randomness Attested by Blockchain," in *Proceedings - 2020 IEEE International Conference on Blockchain, Blockchain 2020*, Institute of Electrical and Electronics Engineers Inc., Nov. 2020, pp. 442–449. doi: 10.1109/Blockchain50366.2020.00064.
- [33] N. Sohrabi and Z. Tari, "Zyconchain: A scalable blockchain for general applications," *IEEE Access*, vol. 8, pp. 158893–158910, 2020, doi: 10.1109/ACCESS.2020.3020319.
- [34] A. Ajorlou and A. Abbasfar, "An Optimized Structure of State Channel Network to Improve Scalability of Blockchain Algorithms," in *2020 17th International ISC Conference on Information Security and Cryptology (ISCISC)*, 2020, pp. 73–76. doi: 10.1109/ISCISC51277.2020.9261916.
- [35] T. Miyamae et al., "ZGridBC: Zero-knowledge proof based scalable and private blockchain platform for smart grid," *IEEE International Conference on Blockchain and Cryptocurrency, ICBC 2021*, pp. 2021–2023, 2021, doi: 10.1109/ICBC51069.2021.9461122.
- [36] C. P. Soi, S. Delgado-Segura, J. Herrera-Joancomartí, and G. Navarro-Arribas, "Analysis of the SegWit adoption in Bitcoin," 2019. [Online]. Available: <https://github.com/bitcoin/bitcoin/blob/a6a860796a44a2805a58391a009ba22752f64e32/src/consensus/consensus.h#L9>
- [37] M. Misbhauddin, A. Alabdulatheam, M. Aloufi, H. Al-Hajji, and A. Alghuwainem, "MedAccess: A Scalable Architecture for Blockchain-based Health Record Management," *2020 2nd International Conference on Computer and Information Sciences, ICCIS 2020*, 2020, doi: 10.1109/ICCIS49240.2020.9257720.
- [38] A. Kancharla, J. Seol, N. Park, and H. Kim, "Slim chain and dependability," *BSCI 2020 - Proceedings of the 2nd ACM International Symposium on Blockchain and Secure Critical Infrastructure, Co-located with AsiaCCS 2020*, pp. 180–185, 2020, doi: 10.1145/3384943.3409435.

- [39] S. H. Sohan, M. Mahmud, M. A. B. Sikder, F. S. Hossain, and R. Hasan, "Increasing Throughput and Reducing Storage Bloating Problem Using IPFS and Dual-Blockchain Method," *International Conference on Robotics, Electrical and Signal Processing Techniques*, pp. 732–736, 2021, doi: 10.1109/ICREST51555.2021.9331254.
- [40] G. Wang, Z. J. Shi, M. Nixon, and S. Han, "SMChain: A scalable blockchain protocol for secure metering systems in distributed industrial plants," *IoTDI 2019 - Proceedings of the 2019 Internet of Things Design and Implementation*, pp. 249–254, 2019, doi: 10.1145/3302505.3310086.
- [41] G. Del Monte, Di. Pennino, and M. Pizzonia, "Scaling blockchains without giving up decentralization and security: A solution to the blockchain scalability trilemma," in *CRYBLOCK 2020 - Proceedings of the 3rd Workshop on Cryptocurrencies and Blockchains for Distributed Systems*, Part of *MobiCom 2020*, Association for Computing Machinery, Sep. 2020, pp. 71–76. doi: 10.1145/3410699.3413800.
- [42] C. Ehmke, F. Wessling, and C. M. Friedrich, "Proof-of-property: A lightweight and scalable blockchain protocol," *Proceedings - International Conference on Software Engineering*, no. January, pp. 48–51, 2018, doi: 10.1145/3194113.3194122.
- [43] G. He, W. Su, and S. Gao, "Chameleon: A Scalable and Adaptive Permissioned Blockchain Architecture," *Proceedings of 2018 1st IEEE International Conference on Hot Information-Centric Networking, HotICN 2018*, pp. 87–93, 2019, doi: 10.1109/HOTICN.2018.8606007.
- [44] Q. Le-Dang and T. Le-Ngoc, "Scalable Blockchain-based Architecture for Massive IoT Reconfiguration; Scalable Blockchain-based Architecture for Massive IoT Reconfiguration," 2019.
- [45] Y. Hassanzadeh-nazarabadi, "LightChain: Scalable DHT-Based Blockchain," vol. 32, no. 10, pp. 2582–2593, 2021.
- [46] S. Linoy, H. Mahdikhani, S. Ray, R. Lu, N. Stakhanova, and A. Ghorbani, "Scalable privacy-preserving query processing over ethereum blockchain," in *Proceedings - 2019 2nd IEEE International Conference on Blockchain, Blockchain 2019*, Institute of Electrical and Electronics Engineers Inc., Jul. 2019, pp. 398–404. doi: 10.1109/Blockchain.2019.00061.
- [47] J. Xu, S. Wang, A. Zhou, and F. Yang, "Edgence: A Blockchain-Enabled Edge-Computing Platform for Intelligent IoT-Based dApps," no. April, pp. 78–87, 2020.
- [48] M. Bez, G. Fornari, and T. Vardanega, "The scalability challenge of ethereum: An initial quantitative analysis," *Proceedings - 13th IEEE International Conference on Service-Oriented System Engineering, SOSE 2019, 10th International Workshop on Joint Cloud Computing, JCC 2019 and 2019 IEEE International Workshop on Cloud Computing in Robotic Systems, CCRS 2019*, pp. 167–176, 2019, doi: 10.1109/SOSE.2019.00031.
- [49] K. Azbeg, O. Ouchetto, and S. Jai Andaloussi, "BlockMedCare: A healthcare system based on IoT, Blockchain and IPFS for data management security," *Egyptian Informatics Journal*, vol. 23, no. 2, pp. 329–343, 2022, doi: 10.1016/j.eij.2022.02.004.
- [50] E. Bandara, X. Liang, P. Foytik, S. Shetty, N. Ranasinghe, and K. De Zoysa, "Rahasak—Scalable blockchain architecture for enterprise applications," *Journal of Systems Architecture*, vol. 116, no. February, p. 102061, 2021, doi: 10.1016/j.sysarc.2021.102061.
- [51] C. Lin, N. Ma, X. Wang, and J. Chen, "Rapido: Scaling blockchain with multi-path payment channels," *Neurocomputing*, vol. 406, pp. 322–332, Sep. 2020, doi: 10.1016/j.neucom.2019.09.114.
- [52] J. Jayabalan and N. Jeyanthi, "Scalable blockchain model using off-chain IPFS storage for healthcare data security and privacy," *J Parallel Distrib Comput*, vol. 164, pp. 152–167, 2022, doi: 10.1016/j.jpdc.2022.03.009.

# Machine Learning Techniques for Diabetes Classification: A Comparative Study

Hiri Mustafa<sup>1</sup>, Chrayah Mohamed<sup>2</sup>, Ourdani Nabil<sup>3</sup>, Aknin Noura<sup>4</sup>

FS, Abdelmalek Essaadi University, TIMS LABORATORY, Tetuan, Morocco<sup>1,3,4</sup>  
ENSATE, Abdelmalek Essaadi University, TIMS LABORATORY, Tetuan, Morocco<sup>2</sup>

**Abstract**—In light of the growing global diabetes epidemic, there is a pressing need for enhanced diagnostic tools and methods. Enter machine learning, which, with its data-driven predictive capabilities, can serve as a powerful ally in the battle against this chronic condition. This research took advantage of the Pima Indians Diabetes Data Set, which captures diverse patient information, both diabetic and non-diabetic. Leveraging this dataset, we undertook a rigorous comparative assessment of six dominant machine learning algorithms, specifically: Support Vector Machine, Artificial Neural Networks, Decision Tree, Random Forest, Logistic Regression, and Naive Bayes. Aiming for precision, we introduced principal component analysis to the workflow, enabling strategic dimensionality reduction and thus spotlighting the most salient data features. Upon completion of our analysis, it became evident that the Random Forest algorithm stood out, achieving an exemplary accuracy rate of 98.6% when 'BP' and 'SKIN' attributes were set aside. This discovery prompts a crucial discussion: not all data attributes weigh equally in their predictive value, and a discerning approach to feature selection can significantly optimize outcomes. Concluding, this study underscores the potential and efficiency of machine learning in diabetes diagnosis. With Random Forest leading the pack in accuracy, there's a compelling case to further embed such computational techniques in healthcare diagnostics, ushering in an era of enhanced patient care.

**Keywords**—Machine learning; support vector machine; artificial neural networks; decision tree; random forest; logistic regression; Naive Bayes; principal component analysis; classification; diabetes

## I. INTRODUCTION

In recent times, diabetes has prominently risen as a pervasive and potentially lethal ailment, with its effects resonating across age groups and genders. This condition, fundamentally shaped by the body's compromised insulin production, interferes with carbohydrate metabolism. This interference results in heightened blood sugar levels, precipitating a slew of symptoms such as augmented thirst, hunger, and frequent urination [1]. A concerning facet of this disease is its accentuated and adverse impact on women, as reflected in their lower survival rates and compromised quality of life [2].

The malaise manifests in three main forms: Type 1, Type 2, and gestational diabetes. Type 1 is predominantly an autoimmune disorder seen in children, leading to the annihilation of pancreatic insulin-producing cells. In contrast, Type 2 emerges when there's heightened insulin resistance

across various organs, eventually pushing the pancreas beyond its production capacities. An added layer of complexity is gestational diabetes, which particularly afflicts pregnant women owing to their pancreas's insufficient insulin output during pregnancy [2]. Furthermore, the diabetes spectrum has more grim facets, capable of inducing long-term harm and malfunctioning in diverse organs like the eyes, kidneys, heart, blood vessels, and nerves [4].

Given the multifarious nature of this disease, physicians find themselves navigating a diagnostic labyrinth. Early diagnosis becomes paramount, serving as the linchpin in circumventing and mitigating potential complications [5]. Fortunately, recent technological strides, predominantly within the machine learning spectrum, proffer novel solutions. Machine learning, a potent sub-discipline of artificial intelligence, harnesses algorithms and statistical frameworks to parse voluminous datasets, unveiling patterns and correlations that often remain concealed from conventional statistical techniques [3].

Positioned against this backdrop, our study delves into the potential of machine learning as a transformative tool in diabetes diagnostics. Six pivotal machine learning classification paradigms - namely, Support Vector Machine, Artificial Neural Networks, Decision Tree, Random Forest, logistic regression, and Naive Bayes - are meticulously examined using the PIDD dataset. By anchoring our assessment on accuracy, we render a holistic comparison of these algorithms' performance nuances.

The paper is structured to facilitate a coherent reader journey. Post this introduction in Section I, Section II immerses into the expansive realm of related works, detailing classification modalities used previously in diabetes prediction. Section III sheds light on our chosen methodologies and intricacies of the PIDD dataset. The crux of our findings unfolds in Section IV, with Section V diving into discussions and implications of these outcomes. Finally, Section VI encapsulates our conclusions, while also hinting at prospective research trajectories.

As we traverse this research landscape, our study is guided by the pressing questions: How do these machine learning paradigms stack against each other for diabetes prediction on the PIDD dataset? Furthermore, can they truly emerge as reliable instruments for diabetes diagnostics?

## II. RELATED WORK

In the annals of modern healthcare research, the strategic deployment of machine learning to grapple with the monumental challenge of diabetes classification has unfailingly occupied a spotlight [6]. The intrigue and allure of this intersection between computational prowess and medical insight have galvanized countless researchers to charter previously unexplored terrains.

Vandana Bavkar, with an academic rigor that's now cited extensively, delivered a magnum opus—a systematic review that scrutinized the versatile applications of machine learning, data mining techniques, and tools in the expansive canvas of diabetes research [6]. His explorations weren't just confined to the realms of prediction and diagnosis. They ventured further, diving deep into the intricacies of diabetic complications, the mystique of genetic predispositions juxtaposed against environmental triggers, and the labyrinth of healthcare management. It was in the revelations of Bavkar's investigation that the bedrock importance of prediction and diagnosis was underscored, positioning them as cornerstone applications of machine learning in the diabetes research tapestry [7].

Parallel to Bavkar's seminal work, Hassan et al. [8] charted their own research trajectory, focusing on the prediction dynamics of diabetes mellitus. Armed with a range of machine-learning classifiers, their study served as a testing ground for techniques such as K-nearest neighbors, Support Vector Machine, and Decision Tree. The metrics they employed—precision, accuracy, sensitivity, and specificity—offered a comprehensive lens through which to evaluate the performance of these classifiers.

Further enriching this research milieu, Kaur et al. delineated a study wherein a quintet of predictive models was brought to the fore [9]. These included stalwarts like Decision Tree, Support Vector Machine, and Naive Bayes. The Pima Indian Diabetes dataset and the R Data Manipulation Tool became their canvas. In a different vein, Zhang et al. concocted a rather innovative approach, introducing a hybrid model that synergized random K-means with Decision Tree, specifically tailored to forecast diabetes risk [10]. Other scholarly forays in this domain have seen the inception of predictive architectures grounded on the Weighted Feature Selection of Random Forest and the XGBoost Ensemble Classifier [11]. Yet another groundbreaking initiative leaned into a logistic regression model, ingeniously augmented by the feature transformation capabilities of XGBoost [12].

Each of these studies, while diverse in methodology and focus, echoes a singular sentiment: the paramount importance of machine learning's role in not just predicting and classifying diabetes, but also in unearthing the intricate dance of genetics and environment, and in revolutionizing healthcare delivery for diabetic patients.

But herein lies an undeniable truth. Despite the richness of insights and the plethora of methodologies that have emerged from these academic odysseys, the horizon of diabetes classification using machine learning still holds vast expanses yet to be charted. The quest for impeccable prediction accuracy

continues, as does the endeavor to spotlight risk factors in their nascent stages. With this study, our ambition is lucidly clear: to augment the extant knowledge reservoir by meticulously assessing the efficacy of a spectrum of machine learning algorithms, all in the context of the revered Pima Indians Diabetes Data Set [13].

To punctuate our intentions and situate our efforts in the grander scheme of academic pursuits, it's paramount to acknowledge the teeming body of studies that have previously addressed this challenge. This dense and rich academic tapestry underscores both the significance and the complexity of the diabetes classification conundrum.

## III. DATASET AND METHODS

### A. Dataset Description

In this study, we utilized the Pima Indians Diabetes Data Set [14], which is a widely used dataset in diabetes research. This dataset was originally collected by the National Institute of Diabetes and Digestive and Kidney Diseases and is available for public use from the UCI Machine Learning Repository. The dataset consists of 768 instances, each containing information about female patients of Pima Indian heritage. The dataset includes various attributes such as age, BMI, blood pressure, skin thickness, insulin level, and diabetes pedigree function, along with the target variable indicating whether the patient has diabetes or not.

The clinical descriptors for these attributes are presented in Table I.

TABLE I. THE CLINICAL DESCRIPTORS OF THE VARIABLES

Number	Attribute	Description	Type
1	Npreg	Number of pregnancies	Numeric
2	Glu	Plasma glucose concentration	Numeric
3	BP	Diastolic blood pressure (mm Hg)	Numeric
4	SKIN	Triceps skinfold thickness, (mm)	Numeric
5	Insulin	Insulin dose, (mu U/ml)	Numeric
6	BMI	Body Mass Index (weight in kg/ (size m) <sup>2</sup> )	Numeric
7	PED	Diabetes pedigree function (heredity)	Numeric
8	Age	Age (Year).	Numeric
9	class	Target variable (0 or 1)	Numeric

### B. Methods

The intellectual pursuit of understanding diabetes through the prism of machine learning necessitates the application of a robust and multifaceted methodology. In light of this, our investigation unfurled in a series of calibrated steps, each meticulously designed to serve a specific purpose within the broader research framework.

1) *Data preprocessing*: Central to the fabric of any data-driven study is the sanctity of the data itself. Recognizing this, our first port of call was to refine and purify the data landscape. We embarked on a rigorous journey of data preprocessing, which, at its core, was about ensuring the



reliability and accuracy of the outcomes. Recognizing the potential pitfalls of missing values, these were diligently identified and addressed with a strategic blend of imputation or outright deletion, depending on the context.

2) *Feature selection*: Beyond just raw data, the richness of features often dictates the nuances of the results. With this philosophy in mind, we ventured into the realm of feature selection. The objective was straightforward yet critical: to streamline the dataset by spotlighting the most consequential attributes for diabetes classification. From the vast repertoire of available techniques, we leaned on the classical Principal Component Analysis (PCA). It's a tool that elegantly navigates the dimensions of data, projecting from a higher-dimensional space to a lower one, all while retaining features pivotal to dataset variance.

3) *Machine learning algorithms*: With the data landscape prepped, the stage was set to deploy the titans of machine learning. Six algorithms, each renowned for its distinctive virtues and latent challenges, were chosen for the diabetes classification task:

a) *Support Vector Machine (SVM)*: The SVM stands tall as a supervised classifier, renowned for its prowess in both regression and classification tasks [15]. Fig. 1 shows SVM algorithm. Originated by Vapnik [16], SVM's genius lies in its capacity to delineate data into classes, both linearly and non-linearly. At its core, SVM conjures hyperplanes in a high-dimensional milieu. The ultimate aspiration? A hyperplane that segregates data classes with the widest possible margin. Non-linear classification gets a boost through a bouquet of kernel functions, each striving to maximize hyperplane margins [17].

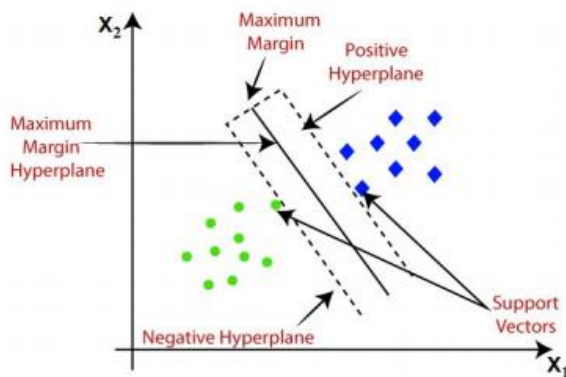


Fig. 1. Support vector machine algorithm [18].

b) *Artificial Neural Networks (ANN)*: Channeling inspirations from the intricate mesh of human neural architecture, ANNs exemplify the confluence of biology and computation [19]. Introduced in the 1950s, ANNs mirror the workings of the human brain's myriad neurons, with artificial neurons and weighted interconnections taking center stage [20]. There are three essential layers in a neural network: input layer, hidden layer, and output layer. The input layer is in charge of accepting data from the user. Fig. 2 shows an example of MLP network with two inputs, five neurons in the

hidden layer. The output layer will provide us with the results. The hidden layer is the layer that sits between the input and output layers. On the same layer, there is no interaction between neurons [21]. If the input vector is  $\vec{x}$ , the weight vector is  $\vec{w}$ , and the activation function is a sigmoid function, the output is as follows:

$$y = \text{sigmoid}(\vec{x} \cdot \vec{w}) \quad (1)$$

and the sigmoid is as follows:

$$\text{sigmoid}(x) = \frac{1}{1+e^{-x}} \quad (2)$$

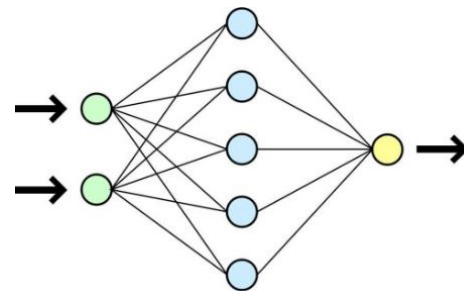


Fig. 2. Example of an MLP network with a hidden layer with two inputs, five neurons in the hidden layer, and one output.

c) *Decision Tree*: Decision Trees, both elegant and insightful, offer a flowchart-like structure to visualize and make decisions. Whether for classification or regression, they rely on a series of attribute tests, guiding data from root to leaf, ultimately culminating in a class prediction [22]. The algorithmic underpinnings encompass three operations: determining terminal nodes, associating non-terminal nodes with tests, and assigning a class to terminal nodes (see Fig. 3) A plethora of algorithms, from ID3 to CTREE, have been proposed for decision tree formulation [23].

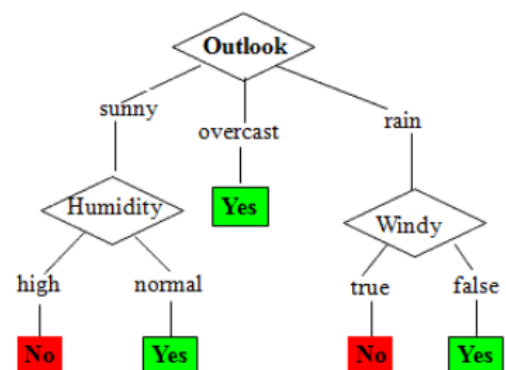


Fig. 3. Example of a decision tree.

d) *Random Forest*: Emerging from the shadows of decision trees is the Random Forest—a brainchild of Breiman [24]. It's an ensemble approach, creating a 'forest' of decision trees from randomly chosen data subsets (see Fig. 4) The collective wisdom of this forest then votes or averages, producing classifications or regressions, respectively [25][26][27].

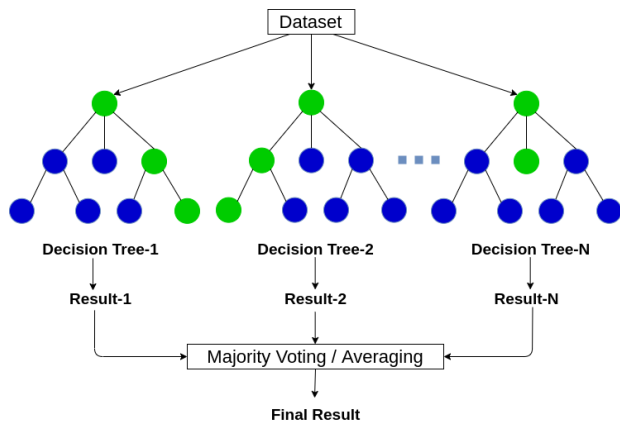


Fig. 4. Random forest.

e) Logistic Regression:

A stalwart in the classification domain, Logistic Regression evaluates probabilities through the sigmoid function, discerning relationships between binary dependent and independent variables. The sigmoid's magic rests in its ability to produce binary outputs based on weighted inputs. If the sigmoid output surpasses 0.5, the prediction is 1; otherwise, it's 0. The sigmoid/logistic function is calculated as follows:

$$y(x) = \frac{1}{1+e^{-x}} \quad (3)$$

where, y is the output which is the result of the weighted sum of the input variables x.

f) Naive Bayes: Grounded in the probabilistic paradigm, the Naive Bayes classifier champions the Bayes Theorem [1]. It presumes that each class feature exists in isolation—hence the "naive" tag. The algorithm computes the posterior probability,

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)} \quad (4)$$

where, P(C|X) is the posterior probability of the target class.

- P(X|C) is the probability of the predictor type.
- P(C) is the probability that class C is correct.
- P(X) is the prior probability of the predictor.

In many intricate real-world scenarios, Naive Bayes has showcased exceptional classification prowess.

IV. EXPERIMENTAL RESULTS

The selection of the Pima Indians Diabetes Data Set for this study was a deliberate choice. This dataset, which has garnered significant attention in the data science community, offers intricate nuances and a wealth of attributes that allow for an exhaustive evaluation of machine learning algorithm performances.

- 1st Experiment: Comprehensive Approach with All Variables.

Our first experiment was anchored in a holistic approach, wherein all available features from the dataset were utilized.

This comprehensive method was designed to create a baseline performance, which future models in our study would either strive to match or surpass. The accuracy metrics corresponding to this experiment, for various algorithms, are tabulated in Table II.

TABLE II. ACCURACY WHEN USING ALL VARIABLES.

Methods	Accuracy validation
Random Forest	0.982
Decision Tree	0.966
SVM	0.954
Logistic Regression	0.794
ANN	0.948
Naïve Bayes Classifier	0.788

As evinced from the results in Table II, the Random Forest classifier emerged as the frontrunner, delivering an impressive accuracy of 0.982, thereby setting a solid benchmark for subsequent experiments.

- 2nd Experiment: Exploring the Power of PCA for Dimensionality Reduction.

Principal Component Analysis (PCA) stands as a testament to the efforts of countless researchers aiming to refine large data volumes into their most significant components. With the aspiration to condense the dataset into its primary four components, representing 71% of its inherent variance, there was an optimistic expectation for data efficiency without sacrificing critical information.

The performance outcomes derived from this approach are detailed in Table III.

TABLE III. ACCURACY WHEN USING THE PCA.

Methods	Accuracy validation
Random Forest	0.97
Decision Tree	0.962
SVM	0.888
Logistic Regression	0.728
ANN	0.826
Naïve Bayes Classifier	0.744

A glance at Table III reveals a pivotal observation: while the Random Forest algorithm continued to exhibit stellar accuracy at 0.97, it was clear that the unmodified data carried nuanced intricacies not entirely captured by PCA. It's a gentle reminder of the delicate balance between data reduction and the preservation of intricate patterns.

- 3rd Experiment: A Deep Dive into Correlation Dynamics.

One of the guiding principles of this experiment was to unearth the relationships and patterns present among the dataset's variables. As machine learning models continue to advance in complexity, a nuanced comprehension of how variables interact and influence each other is paramount.

Fig. 5 and 6 graphically depict the interactions between the class variable and other attributes, as well as the overarching correlation matrix respectively.

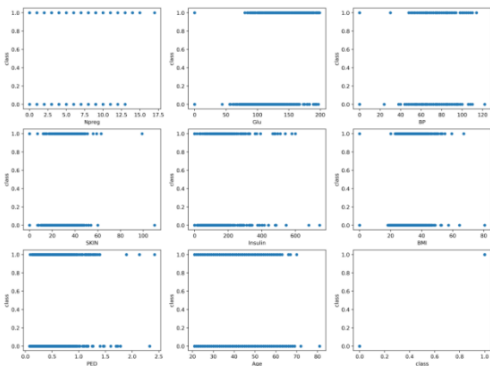


Fig. 5. The class variable as a function of the other variables.

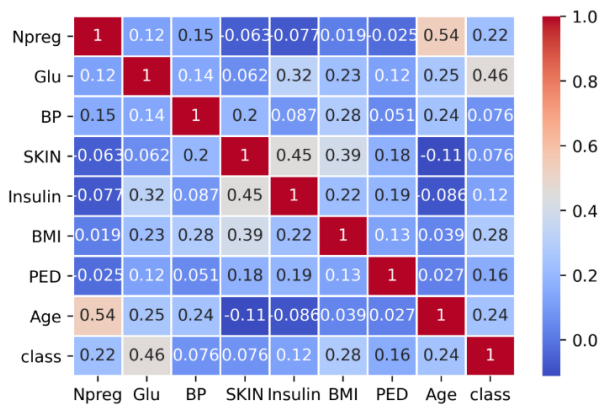


Fig. 6. Correlation matrix.

The metrics arising from this correlation analysis, especially when excluding the 'BP' and 'SKIN' attributes are enumerated in Table IV.

TABLE IV. ACCURACY WITHOUT THE USE OF 'BP' AND 'SKIN'.

Methods	Accuracy validation
Random Forest	0.986
Decision Tree	0.974
SVM	0.964
Logistic Regression	0.792
ANN	0.806
Naïve Bayes Classifier	0.784

An illuminating discovery from this analysis was the marginal contribution of the 'BP' and 'SKIN' variables. By sidelining these variables, the Random Forest algorithm, known for its dynamic adaptability, achieved an apex accuracy of 0.986, highlighting the value of informed feature selection in machine learning.

- 4th Experiment: Spotlight on Prime Features.

The emphasis of this experiment was on identifying and evaluating the predictive power of four critical attributes: number of pregnancies, plasma glucose concentration, body mass index, and age. These features, singled out for their perceived significance, were put to the test to determine their collective predictive prowess.

The outcomes, with focus solely on these attributes are presented in Table V.

TABLE V. ACCURACY WHEN USING THE NUMBER OF PREGNANCIES, PLASMA GLUCOSE CONCENTRATION, BODY MASS INDEX, AND AGE.

Methods	Accuracy validation
Random Forest	0.984
Decision Tree	0.97
SVM	0.964
Logistic Regression	0.788
ANN	0.78
Naïve Bayes Classifier	0.78

While these select attributes showcased substantial predictive capability, the Random Forest algorithm highlighted a noteworthy point: focusing exclusively on them, albeit impactful, didn't outperform its previous benchmarks. The model's accuracy, in this context, peaked at 0.984, subtly reminding us of the intricate dynamics within data.

## V. DISCUSSION

At the confluence of scientific inquiry, we find an unyielding drive towards understanding, clarity, and the quest for tangible insights. Embedded within the heart of this exploration, our study not only aligns with previous findings but also brings forth novel perspectives in the realm of diabetes research [1,15].

One of the standout revelations was the prowess of the Random Forest classifier. Consistent with the observations by Breiman [24] and further corroborated by Liaw and Wiener [27], the Random Forest's consistent performance with the PIMA dataset reaffirms its position of prominence in machine learning applications.

While our experiments were rooted in rigorous methodologies, they were not without their illuminating moments of introspection. Notably, the outcomes from our dimensionality reduction experiment with PCA deviated from what one might expect from theoretical postulations. Such moments, humbling as they are, serve to underline the subtle yet critical chasm that can exist between abstract mathematical formulations and their tangible manifestations in real-world datasets. This deviation nudges us to approach data science with a blend of both rigor and adaptability, being open to unexpected insights.

Diving further into the dataset's granular details, the number of pregnancies, plasma glucose concentration, body mass index, and age have revealed themselves as potential linchpins in diabetes prediction, much in line with previous research findings [5,8]. Yet, the more subtle role of the 'BP' and 'SKIN' attributes reminds us of the broader landscape of attribute interplay and the importance of not viewing any single attribute in isolation.

Conclusively, this exploration has been an enlightening journey, one that reiterates the power of machine learning but equally underscores the necessity for nuanced, iterative data analysis. As the realms of medical diagnostics and data science continue to intersect, it is these intricate dances between data, theory, and application that will pave the way for transformative insights.

## VI. CONCLUSION AND FORWARD PATHWAYS

Throughout our research, we rigorously applied various machine learning methodologies to the PIMA dataset. A consistent standout was the Random Forest classifier, not merely for its algorithmic prowess but for its adaptability and robustness when pitted against intricate datasets like PIMA. The nuanced roles of attributes, especially 'BP' and 'SKIN', underscore the layered complexity within the dataset and the intricacies of diabetes as a medical condition.

Upon deeper examination, it became evident that while some attributes such as the number of pregnancies, plasma glucose concentration, body mass index, and age played pivotal roles in diabetes prediction, others demanded a more careful evaluation. This balance between attribute importance and the broader attribute interplay deepens our understanding and offers a refined perspective on the dataset's potentials and pitfalls.

Looking ahead, there's a wealth of opportunity. The idea of melding deep learning techniques, such as convolutional and recurrent neural networks, with traditional machine learning offers a promising avenue. As medical datasets continue to expand, they will benefit from architectures designed to handle vast amounts of data and extract intricate patterns. This integration could redefine the landscape of medical predictive modeling, particularly for conditions as multifaceted as diabetes. To encapsulate, our findings have been both affirming and enlightening, and the journey ahead in the realms of medical diagnostics and data science is full of promise. Each step we take is more than just academic progression; it is a stride towards enhancing medical prediction and, ultimately, patient outcomes.

## REFERENCES

- [1] S. Deepti and S. D. Singh. "Prediction of Diabetes using Classification Algorithms". *Procedia Computer Science*, 132(), 1578–1585, 2018, doi:10.1016/j.procs.2018.05.122.
- [2] I. Aiswarya, J. S and S. Ronak. "Diagnosis of diabetes using classification mining techniques". *International Journal of Data Mining & Knowledge Management Process (IJDKP)*. Feb. 2015, doi : 10.5121/ijdkp.2015.5101.
- [3] W. Emanuel, D. L. Silvia, C. Eleonora, B. Paola and F. Giovanni. "CamurWeb: a classification software and a large knowledge base for gene expression data of cancer". *BMC Bioinformatics*, 19(S10), 245–256, Oct. 2018, doi:10.1186/s12859-018-2299-7.
- [4] Q. Zou, K. Qu, Y. Luo, D. Yin, Y. Ju, and H. Tang, "Predicting Diabetes Mellitus With Machine Learning Techniques," *Front. Genet.*, vol. 9, no. November, pp. 1–10, 2018, doi: 10.3389/fgene.2018.00515.
- [5] V. V. Vijayan and C. Anjali, "Prediction and Diagnosis of Diabetes Mellitus -A Machine Learning Approach," 2015 IEEE Recent Adv. Intell. Comput. Syst. RAICS 2015, no. December, pp. 122–127, 2016, doi: 10.1109/RAICS.2015.7488400.
- [6] V. C. Bavkar and A. A. Shinde, "Machine learning algorithms for Diabetes prediction and neural network method for blood glucose measurement". *Indian Journal of Science and Technology* 14(10): 869–880, 2021, doi: 10.17485/IJST/v14i10.2187.
- [7] Y. Liu et al., "Machine Learning For Tuning, Selection, And Ensemble Of Multiple Risk Scores For Predicting Type 2 Diabetes," *Risk Management and Healthcare Policy*, Volume 12(), 189–198, Nov. 2019, doi:10.2147/rmhp.s225762.
- [8] F. Hassan and M. E. Shaheen, "Predicting Diabetes from Health-based Streaming Data using Social Media, Machine Learning and Stream Processing Technologies," *International Journal of Engineering Research and Technology*. ISSN 0974-3154, Volume 13, pp. 1957-1967, Number 8. 2020.
- [9] K. Harleen and K. Vinita, "Predictive modelling and analytics for diabetes using a machine learning approach," *Applied Computing and Informatics*, Vol. 18 No. 1/2, pp. 90-100, Mar. 2018, doi:10.1016/j.aci.2018.12.004.
- [10] Z. Hancui, C. Shuyu, C. Wenqian, and W. Tianshu, "A hybrid prediction model for type 2 diabetes using K-means and decision tree," 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), pp. 386–390, Nov. 2017, doi:10.1109/ICSESS.2017.8342938.
- [11] W. Zhiliang and X. Zhongxian, "A Risk Prediction Model for Type 2 Diabetes Based on Weighted Feature Selection of Random Forest and XGBoost Ensemble Classifier," *Eleventh International Conference on Advanced Computational Intelligence (ICACI)*, pp. 278–283, Jun. 2019, doi:10.1109/ICACI.2019.8778622.
- [12] Z. B. Xiangyan et al., "Novel binary logistic regression model based on feature transformation of XGBoost for type 2 Diabetes Mellitus prediction in healthcare systems," *Future Generation Computer Systems*, 129, pp.1-12, Apr. 2022, doi: 10.1016/j.future.2021.11.003.
- [13] O. Adigun, F. Okikiola, N. Yekini, and R. Babatunde, "Classification of Diabetes Types using Machine Learning," *International Journal of Advanced Computer Science and Applications*, Vol. 13, No. 9, 2022.
- [14] "Pima Indians Diabetes Dataset | Kaggle," accessed 06 July 2023.
- [15] N. P. Tigga and S. Garg, "Prediction of Type 2 Diabetes using Machine Learning Classification Methods," *Procedia Comput. Sci.*, vol. 167, pp. 706-716, 2020, doi: 10.1016/j.procs.2020.03.336.
- [16] C. CORINNA and V. VAPNIK, "Support-Vector Networks," *Mach. Learn.*, vol. 20, 1995, pp. 273-297.
- [17] I. Ahmad et al., "Performance Comparison of Support Vector Machine, Random Forest, and Extreme Learning Machine for Intrusion Detection," *IEEE Access*, vol. 6, pp. 33789-33795, 2018, doi: 10.1109/ACCESS.2018.2841987.
- [18] M. M. Abdelsalam and M. A. Zahran, "A Novel Approach of Diabetic Retinopathy Early Detection Based on Multifractal Geometry Analysis for OCTA Macular Images Using Support Vector Machine," *IEEE Access*, vol. 9, pp. 22844-22858, 2021, doi: 10.1109/ACCESS.2021.3054743.
- [19] J. Choi et al., "Convolutional Neural Network Technology in Endoscopic Imaging: Artificial Intelligence for Endoscopy," *Clin. Endosc.*, vol. 53, pp. 117-126, 2020, doi: 10.5946/ce.2020.054.
- [20] B. Alic, L. Gurbeta, and A. Badnjevic, "Machine learning techniques for classification of diabetes and cardiovascular diseases," 6th mediterranean conference on embedded computing (MECO) 2017 Jun 11 (pp. 1-4). *IEEE*. doi:10.1109/MECO.2017.7977152.
- [21] Q. Zou et al., "Predicting Diabetes Mellitus With Machine Learning Techniques," *Front. Genet.*, vol. 9, pp. 1-10, 2018, doi:10.3389/fgene.2018.00515.
- [22] H. Sharma and S. Kumar, "A Survey on Decision Tree Algorithms of Classification in Data Mining," *Int. J. Sci. Res.*, vol. 5, pp. 2094-2097, 2016.
- [23] S. Singh and P. Gupta, "Comparative study ID3, cart and C4. 5 decision tree algorithm: a survey." *International Journal of Advanced Information Science and Technology (IJAIST)*, Vol.3, No.7, July 2014, doi:10.15693/ijaist/2014.v3i7.47-52.
- [24] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, pp. 5-32, 2001, doi:10.1023/a:1010933404324.
- [25] V. F. Rodriguez-Galiano et al., "An assessment of the effectiveness of a random forest classifier for land-cover classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 67, pp. 93-104, 2012, doi:10.1016/j.isprsjprs.2011.11.002.
- [26] V. Svetnik et al., "Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling," *J. Chem. Inf. Comput. Sci.*, vol. 43, pp. 1947-1958, 2003, doi: 10.1021/ci034160g.
- [27] A. Liaw and M. Wiener, "Classification and Regression by RandomForest," *R News*, vol. 2, 2002, pp. 18-22.

# An Optimized Survival Prediction Method for Kidney Transplant Recipients

Benita Jose Chalissery<sup>1</sup>, V. Asha<sup>2\*</sup>

Department of MCA, St. Francis College, Koramangala, Scholar at Department of MCA,  
New Horizon College of Engineering, Bangalore, India<sup>1</sup>  
Department of MCA, New Horizon College of Engineering, Bangalore, India<sup>2</sup>

**Abstract**—Human organ transplantation is a lifesaving process for many of the patients suffering from end stage diseases. Transplantation surgeons are often confronted with the question of the expected survival prognosis for this expensive and perilous process. The aim of the work is to identify an optimal model for predicting the survival of the recipient based on the available organ. This study identifies important features of the recipient and donor parameters for training the model. The study compares the performance of the Random Survival Forest (RSF), which is a machine learning method, and the Cox Proportional Hazard (CPH) model, which is a statistical model, to identify the more accurate model for survival prediction. Variations of the C-index, Brier score, and cumulative Area Under Curve evaluate the survival models considered. This study suggests that CPH which is a statistical method is a better option for forecasting graft and patient survival for an improved clinical outcome.

**Keywords**—Cox proportional hazard model; random survival forest; C-index; brier score; area under curve; organ transplantation; survival prognosis

## I. INTRODUCTION

Kidney transplantation is the only option for those patients identified that the dialysis is no longer a viable solution. According to Organ Procurement and Transplantation Network (OPTN), while there were 88,901 patients waiting for kidney transplantation in US, only 25,499 transplantations were performed in year 2022 [1]. In India, there are around 2 lakhs kidney patients waiting for transplantation per year. However only 10,000 transplantations are performed in a year [2]. Kidney from deceased donor has proven to be a better source to reduce the waiting time for the transplant recipients. There was a huge leap in the number of transplantations in United States due to increase in deceased organ donation. But the Delayed Graft Function (DGF) continues in an upward trend and occurred in 24% of adult kidney transplants in 2021 [3]. Increased DGF is a concern, as it increases the risk of acute rejection and death [4]. To reduce the risk of DGF by taking precaution in the selection of donor kidneys, minimizing cold ischemia time, and monitoring of the recipient after transplantation [5]. In this post pandemic era, especially as it is very difficult to procure an organ, transplant surgeon has to select an ideal recipient for the available organ. Despite having a variety of technologies and infrastructure, relatively little of it is used in such crucial life-saving procedures. The reason and motivation for selecting this topic for research work is mainly as a result of the lack of transplantable organs. As the

availability of organs is very less, we have to make sure that each organ is allocated to the right recipient who can ensure maximize life expectancy. Computer algorithms which suggest the best match for a better survival prognosis helps to increase the success rate of post-transplantation.

Exploiting the new-age technologies to identify the correct recipient for the available organ helps to achieve a better survival prognosis. Sophisticated method helps to find a correct patient in less amount of time promoting interinstitutional organ transplantation without affecting the preservation time of the deceased organ [6]. Implementation of an organ harvesting and transplantation network using IoT and Blockchain improve the efficacy of the organ allocation system. It also monitors the pathophysiological changes in donors and recipients which help in improving the overall quality of organ transplantation [7].

There are numerous survival prediction algorithms that use statistical or machine learning algorithms which are suitable for respective field of study. Random Survival Forest is a machine learning algorithm which predicts by leveraging an ensemble of multiple decision tree. RSF is the machine learning method used for survival prediction particularly while handling complex, high-dimensional data and when making predictions is the key objective. Cox proportional hazard is the statistical method used for survival prediction particularly when there is censored data and when understanding the impact of covariates is the primary goal. The Cox Proportional Hazards (CPH) model and the Random Survival Forest (RSF) algorithm are both significant in the field of survival analysis and have distinct advantages and applications. Selection between them is usually based on the specific characteristics of the data and research objectives. The objective of this paper is to compare the accuracy of survival prediction done by CPH which is a statistical method with RSF which is a machine learning method.

The following section summarizes the review and key findings of related works. The following section after the literature review explains the details regarding the dataset and the methods used. In the material and method section selects the two most significant algorithms based on the review of papers involving survival prediction. The following sub-sections of evaluate the performance of the statistical and machine learning models. The discrimination assessment of the models is done by the calculation of Concordance-index. Time-dependent Brier score is as an alternative method to assess the calibration, of both the methods. Even though, Random

\*Corresponding Author

Survival Forest and Cox's proportional hazards model were performing equally well in terms of discrimination (c-index) and in terms of calibration (IBS), there is notable difference in terms of time dependent Receiver Operating Characteristic (ROC) curve or the cumulative Area Under Curve obtained. This paper suggests CPH as an optimized survival prediction method for kidney transplant recipients.

## II. LITERATURE REVIEW

Lentine, Krista L et al., [3] reflected in their paper that amid COVID-19 pandemic, the field of kidney transplantation faced both successes and challenges in broader geographic organ distribution. The United States witnessed a record number of kidney transplants, mainly due to the increase in deceased donor kidney donation. However, disparities in access to living donor kidney transplant persist, especially for non-White and publicly insured patients. Delayed graft function (DGF) continues an upward trend and occurred in 24% of adult kidney transplants in 2021. Five-year graft survival for deceased donor transplant was 88.6% versus 80.7% for recipients aged 18-34 years, and 82.1% versus 68.0% for recipients aged 65 years or older. The rate of deceased donor transplants among pediatric candidates recovered in 2021 from a low in 2020.

In this paper, the authors Grant, Shannon et al. evaluate various goodness-of-fit tests for the Cox proportional hazards model with time-varying covariates [8]. The Cox proportional hazards model is used in survival analysis to assess the relationship between covariates and the hazard rate. However, when the covariates are time-varying, traditional goodness-of-fit tests may not perform well. The authors propose and compare several alternative tests to assess the model's fit to the data. The key findings of the paper may include insights into the accuracy and reliability of different goodness-of-fit tests when applied to this particular scenario. This research is valuable for improving the assessment of how well the Cox model fits the data when dealing with covariate changes over time, which is a common occurrence in survival analysis studies.

Spooner, A., Chen, E., Sowmya, A. et al. did a comparative study of, ten machine learning algorithms that can perform survival analysis [9]. In this study performance and stability of high dimensional and heterogeneous clinical data was carried out. The researchers developed new prediction models that incorporated immunological factors, recipient, and donor variables, and compared their performance with conventional models. They analyzed data from 3,117 kidney transplant recipients in a multicenter cohort. The results showed that using a survival decision tree model significantly increased the accuracy of graft survival prediction compared to a conventional decision tree model. The occurrence of acute rejection within the first-year post-transplant found to be associated with a 4.27-fold increase in the risk of graft failure.

Yoo, K.D., Noh, J., Lee, H. et al. in their work discusses the challenges in analyzing data from clinical trials and cohort studies, particularly those related to dementia [10]. Such data is often high-dimensional, censored, and heterogeneous, making traditional statistical methods insufficient. Machine learning models that can predict the time until a patient develops

dementia have become essential in understanding dementia risks. They offer more accurate results when dealing with complex clinical data. The study compares ten machine learning algorithms combined with eight feature selection methods to analyze high-dimensional and heterogeneous clinical data. The models predict survival to dementia using baseline data from two different studies: the Sydney Memory and Ageing Study (MAS) and the Alzheimer's Disease Neuroimaging Initiative (ADNI). The models achieved promising performance values, with a maximum concordance index of 0.82 for MAS and 0.93 for ADNI.

The authors K. Suresh, C. Severn, and D. Ghosh [11] discuss various types of discrete-time survival models, such as the Cox proportional hazards model, the logistic regression model, and parametric models like the Weibull and exponential models. The authors emphasize the potential benefits of using machine learning algorithms for more accurate and robust predictions, while also acknowledging the challenges and complexities involved in these approaches.

## III. METHODS AND MATERIALS

Random Survival Forest is a well-known Machine learning algorithms to explore the time to event, in order to study the survival prognosis. Cox Proportional Hazard is a classic statistical approach used on deidentified medical data which may have a high proportion of censored data. While handling censored observations, it can parallelly predict hazard ratio to investigate the association between covariates and survival time of a patient [12]. The aim of the study is to compare the accuracy of survival prediction of the two methods. Training using same dataset gives a better comparison of performance for both Random Survival Forest (RSF) model and Cox Proportional Hazard (CPH) model.

### A. Dataset

The proposed study, use the dataset from United Network for Organ Sharing (UNOS). Standard Transplant Analysis and Research (STAR) files consist of de-identified patient-level information of the transplant recipients and waiting list applicants. The dataset covers patient information starting from January 10, 1987. For the purpose of research, a request sent to UNOS for the STAR dataset. UNOS allowed downloading of STAR dataset from file server on signing a non-disclosure agreement. The data for each type of transplantation covers various attributes of both recipient and donor including survival timeline information. A subset of about 2000 patient data, were used for the comparison study. Attributes selected for training the model includes five features of the transplantation data, along with one event indicator and another attribute indicating the time to event. The event indicator, PX\_STATUS is labelled in four classes, to represent DEAD, ALIVE, RETX or LOST. Mapping to binary representation, the value 'False' assigned to ALIVE status and value 'True' assigned to remaining three status. The five features selected are age of the recipient (AGE), age of the donor (AGE\_DON), BMI of the recipient (BMI\_CALC), HLA mismatch number between recipient and donor (HLAMIS) and cold ischemia time (COLD\_ISCH\_KI) for the organ. The attribute time to event PTIME is the time interval between the transplantation

date and the date at which the event happened, indicated in number of days.

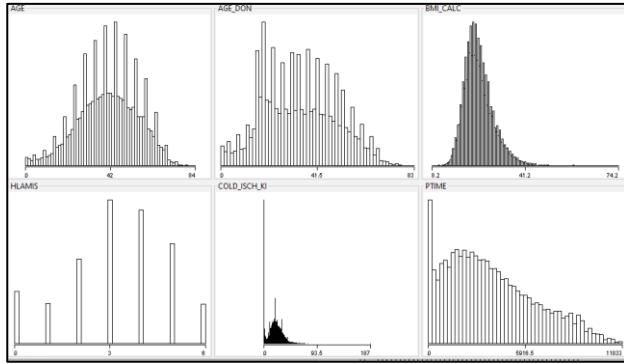


Fig. 1. Feature distribution against time.

Fig. 1 depicts the distribution of selected features against the time span event of patient survival time in days (PTIME).

**B. Staistical Method based Analysis**

Cox proportional-hazards (CPH) model is the statistical method to analyze the risk of several features towards the time to event. This method measures the hazard ratio of covariates on the survival of an individual. Hazard function  $h(t)$  or instantaneous failure rate shows the risk of an event occurring for an individual at any point of time [13]. In case of an individual who has undergone transplantation, the event can be death or re-transplantation at time  $t$ . Calculation of Hazard function  $h(t)$  is as follows [12]:

$$H(t) = H_0(t) \times \exp(b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_kx_k)$$

where  $H_0(t)$  is the cumulative baseline hazard function and  $x_1, x_2, \dots, x_k$  are the subset of predictor variables considered.

The calculation of survival function  $S(t)$ , using CPH model is as follows [12]:

$$S(t) = \exp(-H_0(t) \times PI)$$

where PI, the Prognostic Index. Calculation of PI is as follows:

$$PI = x_1b_1 + x_2b_2 + x_3b_3 + \dots + x_kb_k$$

Survival times are subject to right-censoring. Therefore, we need to consider an individual’s event indicator (PX\_STAT) in addition to survival time (PTIME) [14]. CoxPHSurvivalAnalysis is the python library which is fully compatible to do the required statistical analysis on the dataset and hence used in current analysis of data. PX\_STATUS and PTIME are stored as a structured array. The first field is an indication of observed survival status. Occurrence of event indicated as, ‘True’ value, and ‘False’ value to indicate the remaining status. The second field denoting the observed survival time (PTIME), which corresponds to the number days between the transplantation date and the time of death (if PX\_STATUS == ‘True’) or person contacted last time (if PX\_STATUS == ‘False’).

Cox proportional-hazards model estimates the hazard ratio of a covariate and the effect on the survival of the patient. The extracted hazard ratio and specific distribution generates the

survival time of a patient. Features are used to predict the survival time of an individual. The method overcomes the disadvantage of directly estimating survival time from censored data.

**C. Machine Learning Method based Survival Analysis**

Random survival forests, is an ensemble tree method for analysis of right-censored survival data. Predictions using Random Survival Forest predictions are an aggregation of the predictions of individual trees in the ensemble. Aggregation of the tree-based Nelson-Aalen estimators leads to the construction of the ensemble in Random Survival Forest [15]. The ensemble survival function from random survival forest is as follows:

$$\hat{S}^{rsf}(t|x) = \exp\left(-\frac{1}{B} \sum_{b=1}^B \hat{H}_b(t|x)\right)$$

Corresponding to covariate value  $x$ ,  $\tilde{N}_b^*(s, x)$  is the count of the uncensored events until time  $s$  and  $\tilde{Y}_b^*(s, x)$  is the number of risks at time  $s$ . The estimated conditional cumulative hazard function in each terminal node of a tree using the Nelson-Aalen estimator is as follows:

$$\hat{H}_b(t|x) = \int_0^t \frac{\tilde{N}_b^*(ds, x)}{\tilde{y}_b^*(s, x)}$$

RandomSurvivalForest is the python library used for RSF model creation.

**D. Performance Evaluation of Stastical and Machine Learning Models**

Sample data of six real-world clinical datasets from UNOS evaluates the performance of Cox proportional-hazards model and Random survival forests. Table I shows the clinical dataset used for the analysis. The discriminatory power of five features used to evaluate the predictions done by these predictive models. 20% of training data assess the prediction of the model to predict the survival of a patient after transplantation.

TABLE I. SAMPLE DATA – UNOS DATASET FOR THE FEATURES

	AGE	AGE_DO N	BMI_CAL C	HLAMIS	COLD_ISCH_KI
Sample 1	4	30	21.3	4.0	1.0
Sample 2	10	27	21.3	6.0	9.0
Sample 3	14	5	16.0	5.0	37.0
Sample 4	72	37	22.5	3.0	2.0
Sample 5	72	16	24.1	2.0	17.0
Sample 6	72	62	22.4	5.0	21.0

Graphs generated visualized the evaluation of the results for the given dataset.

a) *Survival Probability Graph*: The survival probability graph in Fig. 2 shows the probability of survival against the number of days, using Cox Proportional Hazard model.

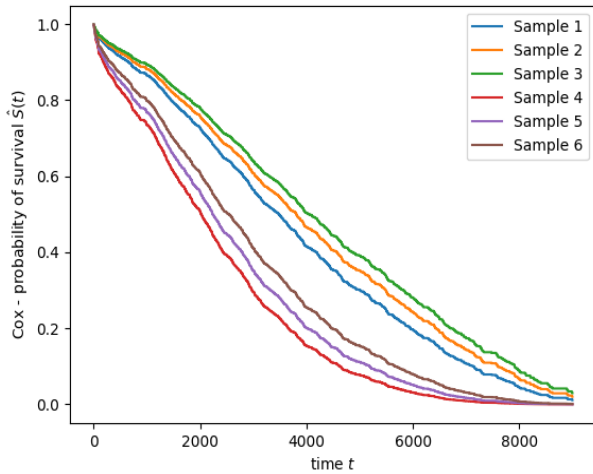


Fig. 2. Probability of survival using Cox PH.

The survival probability graph in Fig. 3 shows the probability of survival against the number of days, using Random Survival Forest model.

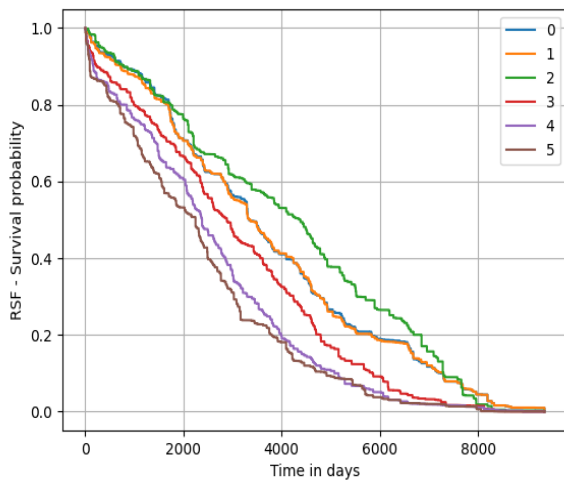


Fig. 3. Probability of survival using RSF.

Fig. 2 and Fig. 3 show that younger recipient is having more survival rate in comparison to older recipient. Both CPH and RSF show similar trends.

b) *Hazard Graph*: The graph in Fig. 4 shows the Cumulative hazard function against survival time in days, using Cox Proportional Hazard model. The graph in Fig. 5 shows the Cumulative Hazard function using Random Survival Forest [16].

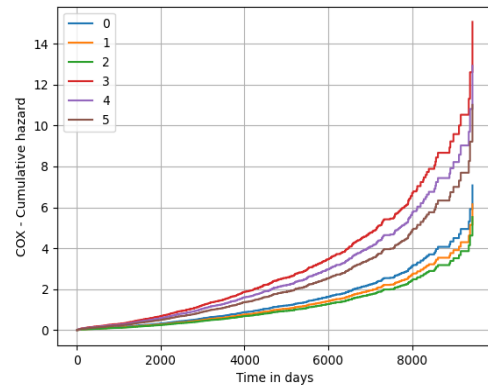


Fig. 4. Cumulative hazard using Cox PH.

Cumulative hazard function reconfirms the finding identified in the survival probability graph.

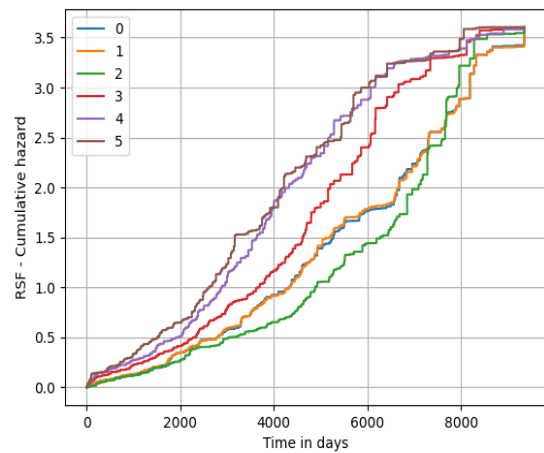


Fig. 5. Cumulative hazard using RSF.

c) *Permutation Importance*: The feature importance of estimators for a given dataset is determined by the permutation importance function. The permutation feature importance measures the increase in the prediction error of the model as a result of permuting the values of a feature. Computation of the permutation importance [17][18],  $i_j$  for the feature  $f_j$ , is as follows:

$$i_j = s - \frac{1}{k} \sum_{k=1}^K s_{k,j}$$

where,  $s$  is the reference score for the data  $D$ , on predictive fitted model  $m$ , calculated for  $K$  repetition for each feature  $f_j$ . The results of permutation feature importance for both the model shows that age of the recipient variable is the main driver of prediction. The variation of data in these columns causes the mean square error in both models to increase. Compared to CPH model, RSF model depends mostly on this variable. Table II shows the permutation feature importance of CPH model and Table III shows the same using RSF model.



TABLE II. PERMUTATION IMPORTANCE-CPH

Permutation Importance-CPH		
	Importance mean	Importance std
AGE	0.024935	0.016718
AGE_DON	0.023710	0.012844
BMI_CALC	0.002368	0.007720
HLAMIS	-0.001456	0.002062
COLD_ISCH_KI	-0.005284	0.007599

TABLE III. PERMUTATION IMPORTANCE-RSF

Permutation Importance-RSF		
	Importance mean	Importance std
AGE	0.049249	0.019136
AGE_DON	0.015333	0.006105
BMI_CALC	0.003339	0.003046
HLAMIS	0.001181	0.004930
COLD_ISCH_KI	0.000656	0.004182

d) *Concordance Index (C-Index)*: C-Index or C-statistic is a measure of predictive accuracy of a model particularly used for survival analysis. C-Index value indicates, a higher risk should result in a shorter time to the adverse event. Therefore, if a model predicts a higher risk score for the first patient ( $\eta_i > \eta_j$ ), we also expect a shorter survival time in comparison with the other patient ( $T_i < T_j$ ).

$$c = \frac{\sum_{i,j} I(T_i > T_j) \cdot I(\eta_j > \eta_i) \cdot \Delta_j}{\sum_{i,j} I(T_i > T_j) \cdot \Delta_j}$$

Split the dataset in training and test sets. Fit the models on the training set. Evaluate the model performances (C-index) on the test set. The desirable values range is between 0.5 and 1. Closer the value towards 1, the more the model differentiates between early events (higher risk) and later occurrences (lower risk). The C-index maintains an implicit dependency on time [19]. The C-index becomes more biased when the amount of censoring is more [20]. CPH gave a Concordance Index of 0.5505 while using RSF, the C-index is calculated as 0.5427. As the number shows CPH gives a better performance than RSF in terms of C-Index.

e) *Brier Score*: Brier Score or Brier Probability score is a measure of the accuracy of the forecast done by a model. The score particularly evaluates the probabilistic prediction. The time-dependent Brier score is an extension of the mean squared error to right censored data. Inverse probability of censoring weights ( $1/\hat{G}(t)$ ) and the model's predicted

probability of upcoming events up to the time t ( $\hat{\Pi}(t|x)$ ), estimates the Brier score as given below [21].

$$BS^{\wedge}(t) = \frac{1}{n} \sum_{i=1}^n \frac{\hat{s}^2(t|x_i)I\{T_i \leq t\}\delta_i}{\hat{G}(T_i - |x_i)} + \frac{(1 - \hat{s}(t|x_i))^2 I\{T_i > t\}}{\hat{G}(t|x_i)}$$

The integrated Brier score at time T is as follows:

$$IBS(T) = \frac{1}{T} \int_0^T BS(t) dt$$

Lower values for the Brier score indicate better prediction performance. Using the Brier score we can calculate the continuous rank probability score (CRPS), defined as the Integrated Brier Score (IBS) divided by time. CPH gave an Integrated Brier score of 0.1958, and 0.1967 while using RSF. In terms of Brier Score, both CPH and RSF are equally performing, however CPH is having slightly better prediction performance.

f) *Receiver Operating Characteristic Curve (ROC)*: Another performance metric to compare the models is time dependent ROC curve [21]. The time-dependent ROC curve is a graphical representation used in survival analysis. This is used to evaluate the performance of predictive models designed to estimate the probability of an event occurring at a specific time in the future. It is also known as the dynamic or cumulative Area Under the Curve (AUC). In the graphical representation of the curve, the x-axis represents time, and the y-axis represents a measure of the model's performance at that specific time. As seen in the Fig. 6, the CPH performance is better than the RSF. The mean value for CPH is 0.602 which is higher than the RSF mean 0.568.

Evaluation of Cumulative hazard function at a time interval of 1000 days calculates the time-dependent risk scores. The plot of CPH shows that the model is doing moderately well on average, with an approximate AUC of 0.602. However, there is a clear difference in prediction performance between the AUC curve of RSF and that of CPH. The performance prediction on the test data increases 15 years after the transplantation surgery. It remains high during the initial 4 to 5 years soon after the surgery and also after 15 years of transplantation. Thus, we can conclude that the model is most effective in predicting death both at the low-term and at high-term using the time-dependent AUC curve.

CPH classify and prioritize parameters using multivariate analysis. The model considered the parameters prioritized for the risk of hazard by Cox Proportional Hazard model. These include CIT, Age of the recipient, HLA mismatch, and BMI calculated. Cox proportional hazard prediction model predict survival days. Prediction accuracy of models evaluated by comparing the predicted survival graph against the actual survival days available as part of the UNOS dataset. Including a fitted model-based prediction in the current allocation policy can enhance the outcome of organ transplantation.

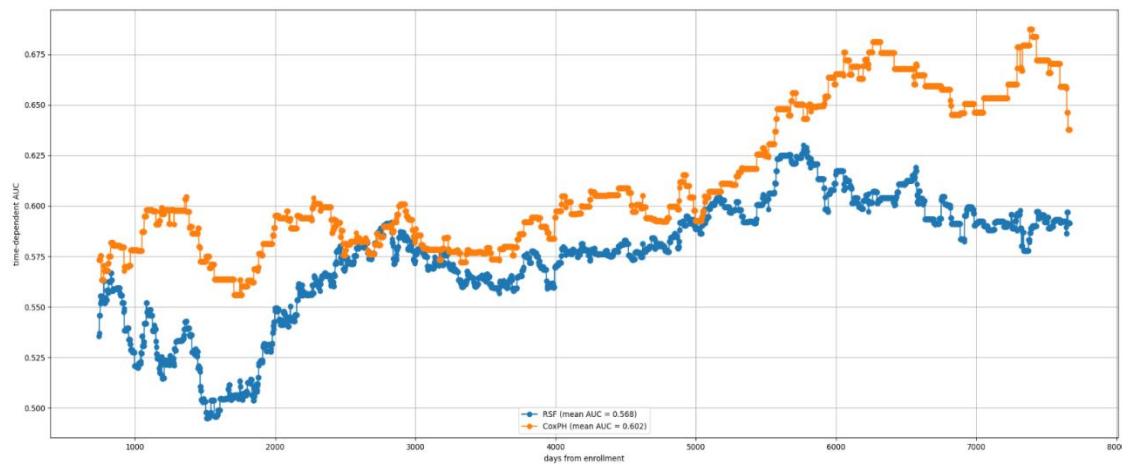


Fig. 6. Comparison of mean AUC -CPH and RSF.

#### IV. RESULTS

Comparing the result with the predictive performance of the Random Survival Forest model, the Cox proportional hazard model performs impartially better on average, mostly due to the better performance in the intervals 4–5 years, and 15–20 years. Even during the period above 5 years, CPH has equally or better performance than RSF. This shows that even though it is convenient to assess overall performance, using mean AUC, even without considering the mean AUC, CPH is a better method to predict the survival prognosis of transplant recipients.

#### V. CONCLUSION

For evaluating survival models considered, variations of the C-index, Permutation Importance, Brier Score, and Cumulative AUC curve proposed over the time are analyzed [21]. The result indicates that both models perform equally well, achieving a concordance index of  $\sim 0.55$ . Evaluation of the prediction of the models is done using alternative methods. Time-dependent Brier score assess the discrimination and calibration, of both the methods. Here again, both the models had the same score of  $\sim 0.196$ . Despite Random Survival Forest and Cox's proportional hazards model performing equally well in terms of discrimination (c-index) and in terms of calibration (IBS), there seems to be a notable difference in terms of time dependent ROC curve or the cumulative AUC. The mean value of AUC with Cox's proportional hazards model outperformed Random Survival Forest. Thus, this paper suggests CPH as an optimized survival prediction method for kidney transplant recipients.

#### REFERENCES

- [1] "Organ donation Statistics | Organdonor.gov," Mar. 01, 2023. Available: <https://www.organdonor.gov/learn/organ-donation-statistics>
- [2] B. Shajan, L. Forrestal, and J. Barret, "Organ shortage Continues to cost Lives," *The Hindu:News*, p. 13, Aug. 6, 2023
- [3] Lentine, Krista L et al. "OPTN/SRTR 2021 Annual Data Report: Kidney." *American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons* vol. 23,2 Suppl 1 (2023): S21-S120. Doi:10.1016/j.ajt.2023.02.004

- [4] M. S. Helfer, J. De Castro Pompeo, O. R. S. Costa, A. R. Vicari, A. M. Ribeiro, and R. C. Manfro, "Long-term effects of delayed graft function duration on function and survival of deceased donor kidney transplants," *Brazilian Journal of Nephrology*, Jun. 01, 2019. [Online]. Available: <https://doi.org/10.1590/2175-8239-jbn-2018-0065>.
- [5] C. Ponticelli, F. Reggiani, and G. Moroni, "Delayed Graft Function in Kidney Transplant: Risk Factors, Consequences and Prevention Strategies," *Journal of Personalized Medicine*, Sep. 21, 2022. [Online]. Available: <https://doi.org/10.3390/jpm12101557>
- [6] B. J. Chalissery, V. Asha, and B. M. Sundaram, "More Accurate Organ Recipient Identification Using Survey Informatics of New Age Technologies," *Atlantis Highlights in Computer Sciences*, Jan. 01, 2021. [Online]. Available: <https://doi.org/10.2991/ahis.k.210913.002>
- [7] B. J. Chalissery and V. Asha, "Blockchain Based System for Human Organ Transplantation Management," *Springer eBooks*, Jan. 01, 2020. [Online]. Available: [https://doi.org/10.1007/978-3-030-41862-5\\_83](https://doi.org/10.1007/978-3-030-41862-5_83)
- [8] Grant, Shannon et al. "Performance of goodness-of-fit tests for the Cox proportional hazards model with time-varying covariates." *Lifetime data analysis* vol. 20,3 (2014): 355-68. Doi:10.1007/s10985-013-9277-1.
- [9] A. Spooner et al., "A comparison of machine learning methods for survival analysis of high-dimensional clinical data for dementia prediction - Scientific Reports," *Nature*, Nov. 23, 2020. [Online]. Available: <https://www.nature.com/articles/s41598-020-77220-w>.
- [10] Yoo, K. D., Noh, J., Lee, H., Kim, Y. S., Lim, C. S., Kim, Y. H., Lee, J. P., Kim, G., & Kim, Y. S. (2017). A machine learning approach using survival statistics to predict graft survival in kidney transplant recipients: a multicenter cohort study. *Scientific Reports*, 7(1). <https://doi.org/10.1038/s41598-017-08008-8>
- [11] K. Suresh, C. Severn, and D. Ghosh, "Survival prediction models: an introduction to discrete-time modeling," *BMC Medical Research Methodology*, vol. 22, no. 1, Jul. 2022, doi: 10.1186/s12874-022-01679-6. Available: <https://doi.org/10.1186/s12874-022-01679-6>
- [12] F. Schoonjans, "Cox regression," *MedCalc*, Aug. 2021, Available: <https://www.medcalc.org/manual/cox-regression.php>
- [13] "The Ultimate Guide to Survival Analysis - Graphpad," *The Ultimate Guide to Survival Analysis - Graphpad*. [Online]. Available: <https://www.graphpad.com/guides/survival-analysis>
- [14] E.-T. Baek et al., "Survival time prediction by integrating cox proportional hazards network and distribution function network - BMC Bioinformatics," *BioMed Central*, Apr. 15, 2021. [Online]. Available: <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-021-04103-w>
- [15] U. B. Mogensen, H. Ishwaran, and T. A. Gerds, "Evaluating random forests for survival analysis using prediction error curves," *Journal of Statistical Software*, vol. 50, no. 11, Jan. 2012, doi: 10.18637/jss.v050.i11. Available: <https://doi.org/10.18637/jss.v050.i11>

- [16] Using Random Survival Forests — scikit-survival 0.21.0. (n.d.). Using Random Survival Forests — Scikit-survival 0.21.0. [https://scikit-survival.readthedocs.io/en/stable/user\\_guide/random-survival-forest.html](https://scikit-survival.readthedocs.io/en/stable/user_guide/random-survival-forest.html)
- [17] “4.2. Permutation feature importance,” Scikit-learn. Available: [https://scikit-learn.org/stable/modules/permutation\\_importance.html](https://scikit-learn.org/stable/modules/permutation_importance.html)
- [18] T. Jensen, “Feature Importance for Any Model using Permutation - Taylor Jensen - Medium,” Medium, Sep. 23, 2022. Available: [https://medium.com/@T\\_Jen/feature-importance-for-any-model-using-permutation-7997b7287aa](https://medium.com/@T_Jen/feature-importance-for-any-model-using-permutation-7997b7287aa)
- [19] Introduction to Survival Analysis with scikit-survival — scikit-survival 0.21.0. (n.d.). Introduction to Survival Analysis With Scikit-survival — Scikit-survival 0.21.0. [https://scikit-survival.readthedocs.io/en/stable/user\\_guide/00-introduction.html](https://scikit-survival.readthedocs.io/en/stable/user_guide/00-introduction.html)
- [20] Albanese, N. C. (2022, June 26). How to Evaluate Survival Analysis Models. Medium. <https://towardsdatascience.com/how-to-evaluate-survival-analysis-models-dd67bc10caae>
- [21] “Evaluating Survival Models — scikit-survival 0.21.0.” Available: [https://scikit-survival.readthedocs.io/en/stable/user\\_guide/evaluating-survival-models.html#Time-dependent-Area-under-the-ROC](https://scikit-survival.readthedocs.io/en/stable/user_guide/evaluating-survival-models.html#Time-dependent-Area-under-the-ROC)

# Analyzing RNA-Seq Gene Expression Data for Cancer Classification Through ML Approach

Abdul Wahid, M Tariq Banday  
Department of Electronics & Inst. Technology  
University of Kashmir, Srinagar, India

**Abstract—Purpose:** Ribonucleic Acid Sequencing (RNA-Seq) is a technique that allows an efficient genome-wide analysis of gene expressions. Such analysis is a strategy for identifying hidden patterns in data, and those related to cancer-specific biomarkers. Prior analyses without samples of different cancer kinds used RNA-Seq data from the same type of cancer as the positive and negative samples. Therefore, different cancer types must be evaluated to uncover differentially expressed genes and perform multiple cancer classifications. **Problem:** Since gene expression reflects both the genetic make-up of an organism and the biochemical activities occurring in tissue and cells, it can be crucial in the early identification of cancer. The aim of this study is to classify the RNA-Sequence data into five different cancer forms, such as LUAD, BRCA, KIRC, LUSC, and UCEC, through an ensemble approach of machine learning algorithms. RNA-Seq data for five different cancer types from the UCI Machine Learning Repository are examined in this research. **Methods:** As a first step, the relevant features of RNA-Seq are extricated using Principal Component Analysis (PCA). Then, the extricated features are given to the ensemble of machine learning classifiers to classify the type of cancer. The ensemble of classifiers is built using Support Vector Machine (SVM), Naive Bayes (NB), and K-Nearest Neighbor (KNN). **Results:** The results demonstrated that the proposed ensemble classifier outperformed the existing machine-learning approaches with an accuracy of 99.59%.

**Keywords—RNA-Sequence; gene expression; feature extraction; voting classifier; ensemble approach**

## I. INTRODUCTION

Cancer is a complex disease characterized by the uncontrolled division and growth of abnormal cells in the body, often forming tumors and potentially spreading to other tissues. When cells behave abnormally and divide abnormally, they can damage neighboring cells and form tumors that can be lethal depending on the circumstances. Early detection and appropriate therapy can reduce the chances of harming other cells. Researchers are working to evolve new systems for preliminary cancer detection and categorization in response to the high cancer mortality rate. However, it is challenging to diagnose cancer early due to the disorganized nature of cancer cells. As a result, RNA-Seq analysis can be instrumental in this case [1]. RNA (Ribonucleic acid) is a molecule that plays a critical role in protein synthesis in cells. RNA is made up of a sequence of four different nucleotide bases: adenine (A), guanine (G), cytosine (C), and uracil (U). RNA sequencing (RNA-Seq) is a powerful technique used to study gene expression by determining the sequence of RNA molecules in a sample. In RNA sequencing, RNA is first isolated from the sample and then converted into complementary DNA (cDNA)

using reverse transcription. Next, the cDNA is sequenced using high-throughput sequencing technologies to generate extensive RNA sequence data. RNA sequential datasets can be used for various purposes, such as studying gene expression, identifying genetic mutations, and developing new disease therapies. These datasets can be generated through various techniques, such as RNA sequencing, microarrays, and hybridization. RNA-Seq is a recent and well-liked method for discovering new transcripts and isoforms by delivering more normalized and less noisy data for prediction and classification purposes. The most crucial role of transcriptome profiling is identifying the differentially expressed genes in the body or finding gene variances at various levels. Using RNA-sequencing, identification and quantification may be done all at one spot. To categorize diseases like breast invasive carcinoma (BRCA), colon adenocarcinoma (COAD), renal chromophobe, etc. RNA-Seq data are freely accessible from many databases [2]. However, many dimensions, complexity, and duplication of features make studying RNA gene expression data particularly challenging. Thus, Machine Learning (ML) and deep learning algorithms can be used to extract features [3], [4] automatically.

Machine Language is a subset of Artificial Intelligence (AI) which is accustomed to identifying underlying patterns in data to identify associations between them [5], [6]. In the age of big data, ML is becoming crucial since it is becoming increasingly difficult for humans to recognize trends and patterns in data to make predictions [7], [8]. ML is thus taking over from humans when identifying and forecasting unseen data to enable informed decision-making. By retrieving features from a database without human input, ML generates predictions. There is a growing use of ML almost everywhere [9]. Its common uses include natural language processing, forecasting, aviation management, and biology to identify protein and RNA sequences [10], [11].

The most crucial aspect of RNA-Seq analyses is differential analysis [12]. Traditional differential analysis techniques often match tumor samples to standard samples of the same tumor kind [13], [14]. However, due to its ignorance of additional tumor forms, such a technology could not distinguish between distinct tumor types [15]. Therefore, conducting an in-depth analysis using RNA-Seq data is necessarily better for understanding the causes of different cancers [16]. Furthermore, most studies attempt to locate genes with differential expression to extract the most pertinent properties. Therefore, developing a strategy that incorporates an understanding of various tumors kinds in the study is essential.

Although RNA-Seq data help detect changes at the gene level, working with RNA-Seq data can be difficult due to its spatial properties [17]. Feature engineering, a technique used to address the challenges of high dimensionality and the relatively small number of samples in gene expression data, is a crucial part of computer approaches for gene expression research. In the current study, gene expression features are extracted in order to overcome the curse of dimensionality and an ensemble of three ML methods for cancer classification using gene expression data have been applied with hard voting strategy. Five tumors of RNA-Seq data are used in this investigation. The current study has applied an ensemble of three ML methods for cancer classification using gene expression data. Five tumors of RNA-Seq data is used in this investigation.

The key contributions of this study are following:

- The proposed framework applies multiple ML models to produce a final ensemble model that is rich in diversity.
- Relevant features extricated from the RNA sequence dataset for cancer prediction.
- RNA Sequence data has been analyzed and visualized to infer knowledge.
- Receiver operating characteristics analysis and state-of-the-art analysis has been done to prove the superiority of the proposed approach.

The remaining paper is organized as follows: The literature relating to the current investigation is discussed in Section II. In Section III, the proposed method is covered. The experimental findings are covered in Section IV, and the article is wrapped up in Section V.

## II. LITERATURE REVIEW

First, to categorise cancer, Sterling Ramroach et al. used various machine learning techniques [18]. A dataset for several cancer kinds was downloaded for their study from the online data portal COSMIC. The machine learning models that were used were support vector machine (SVM), neural networks, K closest neighbour (KNN), and random forest (RF). For various cancer types and primary sites, the authors conducted numerous tests. In contrast to other algorithms, RF distinguished itself by achieving significant classification accuracy and being simple to tune.

The boosting deep cascade forest (BCDForest) deep learning algorithm was presented by Yang Guo et al. as the preference for deep neural networks for categorising the cancer RNA. This strategy was used to publicly available microarray data sets encompassing adenocarcinoma, brain, and colon cancer and RNA-Seq data sets containing BRCA, GBM, pan cancers, and LUNG. Each deep forest in this ensemble methodology worked well in predicting the classification outcomes. First, Cascade forests are built using decision tree-based random forests trained to find relevant characteristics in raw data. Next, this result was placed against state-of-the-art classifiers like SVM, KNN, LR, RF, and the original gforest [18]. The authors claimed that their suggested approach produced more precise results.

Yawen Xiao et al. suggested that the multimodal ensemble technique includes KNN, SVM, DTs, RFs, and Gradient Boosting Decision Trees (GBDTs) [19]. Three different cancers were treated using their suggested approach: LUAD, stomach adenocarcinoma (STAD), and BRCA. This tactic was used to train each classifier individually using the supplied data to produce predictions, which were then used to inform a multimodal ensemble approach using deep learning. This technique predicts cancer more accurately than data produced by a single classifier.

Using Voom, Dincer Goksuluk et al. developed a new range of classifiers termed “voomNSC”, “voomNBLDA”, and “voomPLDA” to classify and assess RNA-Sequencing data. VoomNSC uses the NSC approach in conjunction with voom transformation to create classifiers that are more reliable and accurate [3]. Because VoomDLDA and voomDQDA are not sparse bases, they take advantage of all the model’s properties. The sparse base classifier voomNSC uses only the subset of features in the model. The results showed that voomNSC produced the best outcomes compared to PLDA, NBLDA, and NSC.

Paul Ryvkin et al. provided a brand-new numerical method for CoRAL (classification of RNA by analysis of length) [20]. For this reason, the authors sequenced databases of short RNA sequences. Three trimmed adapter sequences were then applied to the dataset, and a FASTA file was generated after completing numerous pre-processing steps. Next, aligned reads were recorded in SAM files by comparing them to a reference file. A SAM file was then created based on the mismatch rate of the readings. Finally, a BAM file containing the aligned and matched genes was created and delivered to CoRAL. CoRAL categorises various RNA sequence types and draws out salient traits from them. This technique categorises short RNA sequences and gives the user a more significant direction.

Hamid Reza Hassanzadeh et al. suggested a cutting-edge pipeline technique to predict the prognosis of cancer patients [21]. The proposed method used Laplacian Support Vector Machines for semi-supervised learning. This technique predicted the survival of patients with neuroblastoma (NB) and kidney cancer (KIRC). It involved four steps where pre-processing is the first step which includes feature metric storage and data analysis. The second step is feature extraction and then next step removes overfitting problems. Using a generalisation strategy as the final step will enable to assess the precision of each model and determine the weights accordingly. In terms of accuracy, this pipeline method performed better than supervised SVM.

Jiande Wu et al. have suggested using several machine-learning algorithms to detect triple-negative breast cancers [5]. In this study, TCGA data were used to evaluate the gene expression levels of 110 breast cancer samples that were triple-negative with 992 non-triple-negative samples. SVM, KNN, Naive Bayes (NB), and DT were the machine learning classification models that were employed. Due to the enormous dimensions of the data, a further step known as feature selection was carried out before classification to obtain the essential features. The categorisation job had accuracy rates of

90%, 87%, 85%, and 87%, respectively. The results demonstrate that SVM outperformed the other techniques.

GeneQC (gene expression quality control), a machine learning-based technique, was proposed by Adam McDermaid et al. to determine the reliability of expression levels precisely from RNA sequencing datasets [15]. The authors used data from seven plant and animal taxa's RNA sequencing. Three different types of information were entered into GeneQC. A SAM file is read by the first mapping, a reference genome FASTA file by the second, and a species-specific annotation file by the third. GeneQC uses two processes: a Perl script to extract features and an R programme to model the mathematical relationships between those features. GeneQC then categorises the reading alignment category for each Genome.

Yawen Xiao et al. presented a stacked sparse auto-encoder, utilising a semi-supervised deep learning methodology [19]. LUAD, STAD, and BRCA were just a few of the cancer types that this approach predicted. This model integrated supervised classification methods with semi-supervised feature extraction techniques to handle labelled and unlabelled data and extract more precise information for cancer prediction. The results demonstrated that the suggested method gave more accurate prediction results when compared to several cutting-edge machine learning classifiers, including SVM, RF, NN, and auto-encoders. In addition, several studies have considered using technologies, including wireless sensor networks, networks, software-defined networking, and the Internet of Things (IoT) [22].

To find biomarkers in high throughput sequencing, Brian Aevermann et al. suggested combining feature selection and the binary manifestation method of a random forest [23]. The authors' analysis supports this by using the NS-Forest version 2.0. Identifying active cell types and under investigation are two goals for which the most recent iteration of NS-Forest is effective. Their study sent a cell with a clustered gene expression assignment to the RF, from which significant features were gleaned using the Gini index. To overcome unfavourable indicators, genes were further prioritised. The top-ranked genes were then determined using a binary expression score. To adjudicate the least number of features, a criterion based on a decision tree and F-Beta score was employed to investigate various combinations of biomarkers. Finally, the human middle temporal gyrus (MTG) was used in tests to gauge the technique's efficiency [24].

Barbara Pes used the homogenous ensemble approach and applied the selection algorithm to several diversified datasets derived from the original set of records. The author worked on high-dimensional benchmarks from various domains, and this ensemble approach led to a significant gain without any degradation of the predictive performance [25].

Table I tabulates the existing literature on cancer classification with advantages and disadvantages, which pave the way to propose a novel ensemble machine learning technique in this study. Compared to the current cancer

classification approaches, the proposed method is different in the way that the RNA features are extricated using PCA and the type of cancer is classified using the proposed ensemble classifier that reduces the computation complexity as the model is constructed using the extricated features alone.

### III. PROPOSED APPROACH

Most traditional cancer classification systems use a single classification method, relying heavily on a specific classification algorithm for accuracy. The performance of a particular classifier may differ depending on the dataset. Therefore, to increase prediction accuracy, a framework must be developed for combining complementary information from different classifiers. The proposed approach is the hybridization of feature extraction and an ensemble of machine learning classifiers that classify cancer using RNA sequence data. Fig. 1 illustrates the block diagram of the proposed approach. This approach consists of the following modules: feature extraction, data splitting, model selection, and voting ensemble classification.

#### A. Feature Extraction

Feature extraction is an essential step in machine learning. It involves selecting and transforming the most relevant information from the input data to create a set of new, more informative features. This can help improve machine learning algorithms' performance by reducing the data's dimensionality and removing noise or irrelevant information. There are many techniques for feature extraction, including Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Independent Component Analysis (ICA). PCA identifies the directions of maximum variance in the data and projects the data onto a new coordinate system defined by these directions, called principal components. The first principal component is the direction of maximum variance in the data. Each subsequent principal component is orthogonal to the previous components and captures the maximum remaining variance. The data is first standardised by performing PCA to have zero mean and unit variance. Then, the covariance matrix is computed, and its eigenvectors and eigenvalues are calculated. The eigenvectors represent the directions of the principal components, and the eigenvalues represent the amount of variance explained by each component. The data can then be projected onto the principal components by multiplying the original data matrix by eigenvectors [34], [35].

Let the dataset be  $D$  consisting of  $x+1$  dimensions. Ignore the labels such that new dataset become  $x$  dimensional.

The mean for every dimension of the whole dataset is computed as follows:

$$D_{\mu} = \frac{x}{D_{size}} \quad (1)$$

The covariance matrix of the whole dataset is computed as follows:

$$Cov_{mat}(D_A, D_B) = \frac{1}{n} \sum_{i=1}^n (A - \bar{A})(B - \bar{B}) \quad (2)$$

TABLE I. REVIEW OF EXISTING CANCER CLASSIFICATION SYSTEMS

Ref.	Methodology	Used Dataset	Metrics	Advantages	Disadvantages
Goksuluk et al., 2019 [3]	Microarray-based classifiers	Synthetic dataset	Accuracy, sparsity, sensitivity, specificity	User-friendly and simple	Prior knowledge of packages is required
Khalifa et al., 2020 [4]	Optimised deep learning	Tumour gene expression dataset	Precision, recall, F1-score, accuracy	Less complex and requires less time to train	Performance is low
Wu et al., 2021 [5]	SVM, KNN, NB, and DT	Cancer Genome Atlas dataset	Accuracy, recall, specificity, precision, F1-score	Efficient	Complexity is high
Ramroach et al., 2020 [9]	RF and Gradient boosting machine	Cancer Genome Atlas dataset	Accuracy	High performance	Complexity is high
Arowolo et al., 2020 [26]	Ensemble classifier	RNA sequence dataset	Accuracy, sensitivity, specificity, precision, recall, F1-score	Less complex	Low accuracy
Yu et al., 2020 [27]	NB, RF, SVM	RNA sequence dataset	Sensitivity, specificity, accuracy, F1-score, AUC	Complexity is low	Interpretation is low
Garcia-Diaz et al., 2020 [28]	Grouping genetic algorithm	RNA sequence dataset	Standard deviation, accuracy	Computation speed is fast	Incomplete exploration of solution space
Mohammed et al., 2023 [29]	Reinforcement learning	Omics dataset	Accuracy	High processing speed	The optimisation is done partially
Arowolo et al., 2021 [30]	KNN and Decision tree	Western Kenya RNA sequence dataset	Accuracy, sensitivity, specificity, precision, recall, F-score	Less complex	Low accuracy
Arowolo et al., 2021 [31]	Genetic algorithm and Ensemble classification	Anopheles Gambiae dataset	Accuracy, sensitivity, specificity, precision, recall, F-score	High specificity	Works for only small datasets
Ramamurthy et al., 2020 [32]	Deep learning	Synthetic dataset	Recall Jaccard index, dice index, correlation coefficient, specificity, F1-score, computational time	High accuracy	More complex
Mohammed et al., 2021 [33]	Stacking ensemble	Cancer Genome Atlas dataset	Accuracy, F1-score, precision, sensitivity, AUC	High accuracy	Less inference

The eigenvectors and the corresponding eigenvalues are computed as follows:

$$\det(D-\lambda I)=0 \quad (3)$$

A  $d \times k$  dimensional matrix is created by selecting the  $k$  eigenvectors with the most significant eigenvalues after the eigenvectors are sorted in decreasing order. Next, the samples are transformed into the new subspace using the eigenvector matrix, yielding the principal components.

### B. Data Splitting

Training and test sets are created from the primary component data. The test set assesses how well each classification model performed, and the training set is used to create classification models. The total sample size determines the ratio for dividing the data into two portions. For example, 70% of the training set is typically used in research, and the remaining 30% is used as the test set. However, the split ratio can be lowered to 50% when there are fewer samples [36] [37]. Like the last example, this ratio might be raised to 80% or 90% if the total number of samples is high enough. The fundamental idea behind determining the ideal splitting ratio is to select a splitting ratio with a sufficient number of samples in both the training and test sets to generate a trustworthy fitted model and test predictions. The test accuracy is sensitive to unit

misclassifications even though the fitted model is ultimately reliable. In our proposed approach, data has been split on the ratio of 70:30.

### C. Model Selection

Selecting the right classifier for a particular machine-learning task is essential to the modelling process. There are a variety of classifiers to choose from, each with its strengths and weaknesses. The factors to consider when selecting a classifier include the type of problem, dataset size, data complexity, and interpretability and performance metrics. Some commonly used classifiers in machine learning are Logistic Regression, K-Nearest Neighbors, Support Vector Machines, Decision Trees, Random Forests, and Naive-Bayes. It is often a good idea to try multiple classifiers and compare their performance on the given task to determine the best option. The ensembles of multiple classifiers can often perform better than a single classifier. After building these machine learning models, only the top-performing models are considered for proposed ensemble model building.

### D. Ensemble of Classifiers

The ensemble of classifiers is built by combining the advantages of three classifiers such as SVM, NB, and KNN.

- Support Vector Machine

The SVM is mainly used for categorisation due to its excellent accuracy and capacity for managing enormous amounts of data. It is a supervised ML algorithm. The goal of the SVM method is to find a hyper-plane that divides the data set into distinct groups in a suitable way for training sets [38]. Linearly separable data can be divided into two groups by a straight line. A line can separate data that are linearly separable in two dimensions. The function of the line can be represented as follows:

$$y=ax+b \quad (4)$$

The above equation can be re-written as follows by replacing  $x$  with  $x_1$  and  $y$  with  $x_2$ :

$$ax_1-x_2+b=0 \quad (5)$$

If  $x$  and  $w$  are defined as  $x = (x_1, x_2)$  and  $w = (a, -1)$ , then (4) is defined as follows:

$$wx+b=0 \quad (6)$$

It is the equation of the hyperplane, which is derived from two-dimensional vectors. This hyperplane is used to make predictions. For example, cancer is defined as having a point above or on the hyperplane and not having a threshold below the hyperplane.

- Naive Bayes

Naive Bayes (NB) classifiers are scalable because the number of parameters required is linear in the learning process's number of variables (features/predictors). A closed-form expression, which takes linear time, can be evaluated to perform maximum-likelihood training [39]. The classifier is a function that is computed as follows:

$$NB_{cl}=\operatorname{argmax}_{k \in \{1, \dots, K\}} P(C_k \pi_{i=1}^n p(x_i|C_k)) \quad (7)$$

- K Nearest Neighbor

K-Nearest Neighbor (KNN) is a supervised algorithm based on the distance function. The distance function, which assesses the degree of similarity or difference between two

samples, is the basis of this classifier. The Minkowski distance metric is computed as follows:

$$MD(x,z)=\left(\sum_{r=1}^d \|x_r-z_r\|^p\right)^{\frac{1}{p}} \quad (8)$$

With KNN, the function is locally approximated, and all computation is delayed until the function is assessed. Normalising the training data can significantly improve accuracy if the features represent different physical units or sizes because this technique relies on distance for classification. In addition, applying weights to neighbour contributions can help classification and regression because it encourages neighbours closer to one another to contribute more to the average than neighbours farther away. When utilising KNN classification or KNN regression, the neighbours are selected from a group of objects for which the class or object property value is known [40].

#### IV. RESULTS AND DISCUSSION

The experiments were evaluated on an Intel(R) Core(TM) i7-6700 processor with 8 GB of RAM under Windows 10. The proposed approach was implemented in Python using the available machine learning packages. The UCI Machine Learning Repository hosts an RNA sequencing dataset containing gene expression data obtained from RNA sequencing of cancer cells and healthy cells. The gene expression levels are measured for over 20,000 genes, and more than 5,000 samples are in the dataset. The dataset used for experimentation is the RNA sequence dataset. This dataset is from the UCI Machine Learning Repository. The dataset contains information on the gene expression levels of five different cancer forms [41]. They are listed as follows:

- Lung ADenocarcinoma (LUAD)
- BRest invasive CArcinoma (BRCA)
- KIdney Renal Clear cell CArcinoma (KIRC)
- Lung Squamous Cell CArcinoma (LUSC)
- Uterine Corpus Endometrial CArcinoma (UCEC)

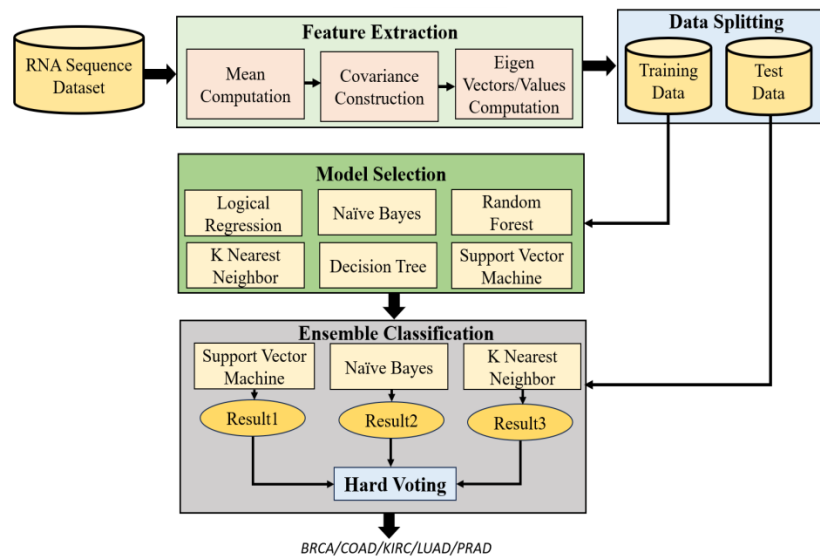


Fig. 1. Block schematic of proposed cancer classification approach.



---

Algorithm I: Cancer RNA Sequence Classification Algorithm

---

**Input:** RNA Sequence Dataset,  $R_D$

**Output:** RNA Cancer type, BRCA/COAD/KIRC/LUAD/PRAD

**Process:**

- 1: for all records in  $R_D$
- 2: Compute mean of RNA sequence data,  $D_\mu$
- 3: Construct covariance matrix,  $Cov_{mat}$
- 4: Calculate Eigen vectors/Eigen values of  $Cov_{mat}$
- 5: Return top  $K$  principal components
- 6: end for
- 7: Traindata, Testdata=split(CancerRNASequencefeatures, label)
- 8: Return Traindata, Testdata
- 9: voting="hard"
- 10: M1=SVM(Traindata, Trainlabel, Testdata)
- 11: M2=NB(Traindata, Trainlabel, Testdata)
- 12: M3=KNN(Traindata, Trainlabel, Testdata)
- 13: VotingEnsembleModel(Traindata, Trainlabel, Testdata)
- 14: hardvotingclassifier=concatenate(M1, M2, M3)
- 15: hardvotingclassifier.fit(Traindata, Trainlabel)
- 16: classification=hardvotingclassifier.predict(Testdata)
- 17: Return RNACancerclass

---

There are 20531 attributes over 801 occurrences. The most dangerous type of cancer for women is BRCA. The most common type of kidney carcinoma, known as KIRC, accounts for 70–80% of instances of the disease and has a high mortality rate globally. LUAD is a common type of cancer. Around 40% of all lung cancer diagnoses are due to it. It primarily attacks non-smokers. LUAD is typically discovered by accident and spreads more slowly than other forms of lung cancer. Smokers are likelier to get LUSC, the second most prevalent lung cancer. Airborne smoke particles often reside in the middle of the lung and transmit LUSC cancer. Undiagnosed in its early stages, UCEC is a recurrent prenatal malignancy. It affects more women than any other type of cancer. Due to the lack of information on its biomarkers for early detection and treatment, it has a high mortality rate. Fig. 2 depicts the distribution of cancer classes.

**E. Principal Component Analysis**

Fig. 3 depicts the scatter plot of principal components. The dimension of the RNA sequence data is high and in order to improve the performance of the classification task, the dimension of the dataset has been reduced and features are extricated using PCA. Experimentation has been done with varying number of principal components and using trial and error approach the number of principal components used in the proposed approach is five. The reason behind the achievement of significant results using five principal components is that the dataset consists of five cancer classes. It is observed from the scatter plot that there are similarities in LUAD, BRCA, and COAD cancer classes. The KIRC and PRAD are scattered separately as there are dissimilarities exist in these classes compared to LUAD, BRCA, and COAD.

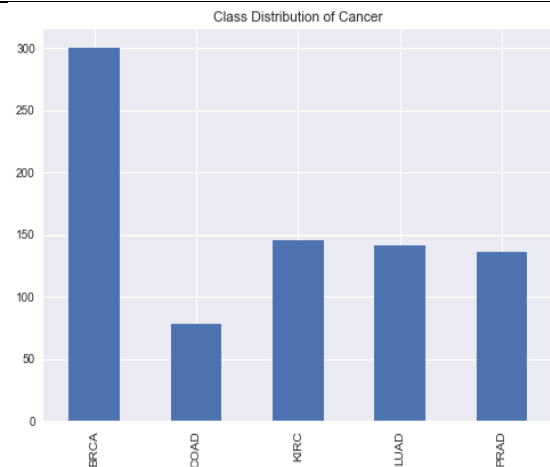


Fig. 2. Distribution of cancer classes.

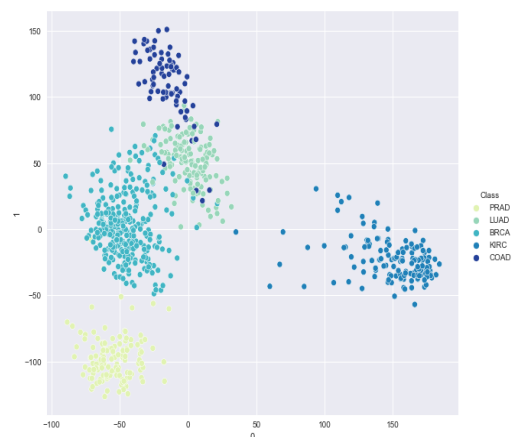


Fig. 3. Scatter plot of principal components.

### F. Performance Evaluation

Performance evaluation in machine learning is assessing the accuracy and effectiveness of a trained model. It is essential to evaluate the performance of a machine learning model to determine its effectiveness in solving a specific problem [42]. The model's performance can be improved by tuning the hyper-parameters. Various metrics for evaluating a model's performance include Accuracy, Confusion Matrix, Precision, Recall, F1-Score, AUC (Area- Under-the-Curve)-ROC.

Fig. 4 depicts the confusion matrix for the classification of cancer RNA sequences. There are  $n$  columns and  $n$  rows in a confusion matrix, where each column represents a predicted classification, and each row represents the true classification [43]. To determine the model's accuracy, it is possible to examine the values along the diagonal - a good model will have a high diagonal value and low values off it. Furthermore, one can determine where the model is having difficulty by examining the highest values, not on the diagonal. These analyses help identify cases where the model's accuracy is high but consistently misclassifies the same data.

A classification report is a technique used to evaluate the performance of machine learning models in multiclass classification problems. It comprehensively summarises the model's performance on various evaluation metrics such as precision, recall, F1-score, and support. Fig. 5 depicts the classification report with the considered performance metrics. The precision, recall, f1-score, and support are computed for all the cancer classes. Furthermore, the macro average and weighted average are also computed to know the performance of the studied cancer ensemble classifier. The accuracy obtained is approximately 100% using the proposed ensemble approach for classifying the cancer RNA sequences.

Table II compares training and testing scores of the existing and proposed cancer classifications. It is seen that the proposed approach performed significantly well in training, but the performance is not significant in terms of testing compared to the proposed hybrid ensemble approach.

TABLE II. TRAINING AND TESTING SCORE ANALYSIS

Model	Training Score (%)	Testing Score (%)
LR	99.46	98.59
NB	98.75	99.17
RF	99.46	98.76
KNN	99.46	98.75
DT	98.75	97.51
<b>Proposed</b>	<b>99.64</b>	<b>99.59</b>

### G. ROC Analysis

The ROC (Receiver Operating Characteristic) curve is a graphical representation of the performance of the classifier, showing the trade-off between sensitivity (true positive rate)

and specificity (true negative rate) at different classification thresholds [44], [45]. To create a ROC curve, the classifier is applied to a dataset with known outcomes (i.e., a labelled dataset), and the true positive rate (TPR) and false positive rate (FPR) are calculated for different classification thresholds. The TPR is the proportion of true positive predictions among all positive cases in the dataset, and the FPR is the proportion of false positive predictions among all negative cases in the dataset. These rates are plotted on the y-axis and x-axis for different thresholds, resulting in a curve that starts at the origin (TPR=0, FPR=0) and ends at (TPR=1, FPR=1). The area under the ROC curve (AUC) is a standard metric summarising the classifier's overall performance. For example, an AUC of 0.5 indicates random performance, while an AUC of 1 indicates perfect performance. A higher AUC value indicates better classifier performance distinguishing between the positive and negative classes.

The One-vs-Rest (OvR) classifier and the One-vs-One (OvO) classifier are two common approaches for multiclass classification problems [46]. In the OvR approach, a separate binary classifier is trained for each class, which distinguishes that class from all the other classes. In contrast, the OvO approach trains a binary classifier for each pair of classes. Both approaches can be used to generate ROC curves for multiclass classification problems. In the case of OvR, the ROC curve is generated by computing the false positive rate (FPR) and true positive rate (TPR) for each class's binary classifier. The overall ROC curve is then obtained by combining the individual curves for each class. In the case of OvO, the ROC curve is generated by comparing the predicted class probabilities for each pair of classes and computing the FPR and TPR based on the number of correct and incorrect predictions for each pair.

Finally, the overall ROC curve is obtained by combining the FPR and TPR values for all the pairs of classes. When applied to gene selection methods, OvR and OvO can help to improve the results by reducing the number of false positives and false negatives in the classification process. By treating each class as a separate binary classification problem or training separate models for each pair of classes, OvR and OvO can help to better capture the subtle differences between the different classes, leading to more accurate classification results.

Fig. 6 depicts the ROC analysis for One-vs-Rest (OvR) classifier. The AUC is high for the proposed approach compared to the existing approaches such as LR, NB, RF, and KNN. The reason behind the high performance of the proposed approach is that the extracted features are used for classification. Furthermore, the advantages of the existing classifiers are combined to build the ensemble classifier.

Fig. 7 depicts the ROC analysis for One-vs-One (OvO) classifier. All the plots depict high AUC except COAD versus LUAD, as these cancer classes have high similarity by which classifier cannot differentiate these two classes efficiently.

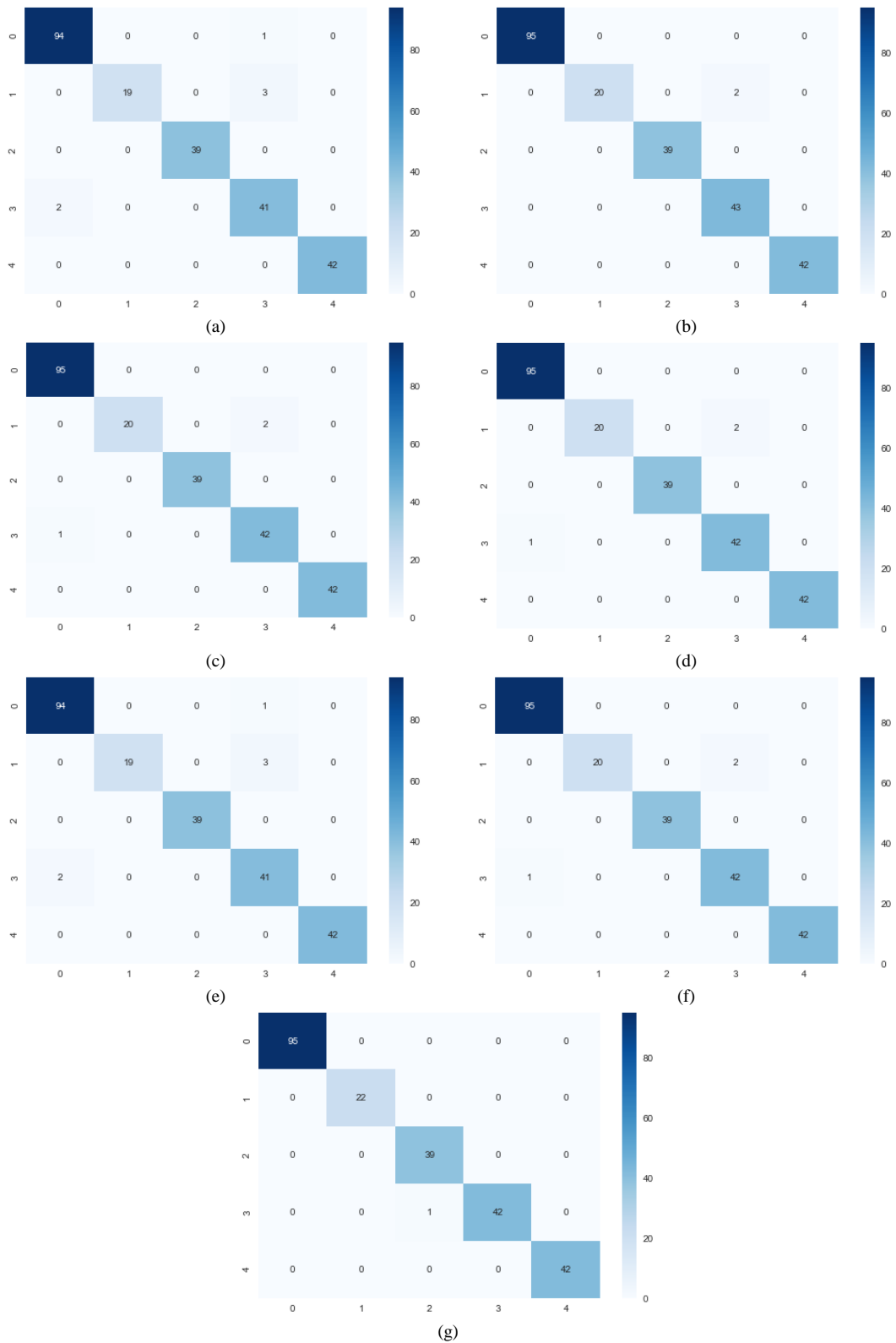


Fig. 4. Confusion matrix (a) Logistic regression (b) Naive Bayes (c) Random forest (d) K nearest neighbor (e) Decision tree (f) Support vector machine (g) Proposed approach.

	precision	recall	f1-score	support		precision	recall	f1-score	support
BRCA	0.98	0.99	0.98	95	BRCA	1.00	1.00	1.00	95
COAD	1.00	0.86	0.93	22	COAD	1.00	0.91	0.95	22
KIRC	1.00	1.00	1.00	39	KIRC	1.00	1.00	1.00	39
LUAD	0.91	0.95	0.93	43	LUAD	0.96	1.00	0.98	43
PRAD	1.00	1.00	1.00	42	PRAD	1.00	1.00	1.00	42
accuracy			0.98	241	accuracy			0.99	241
macro avg	0.98	0.96	0.97	241	macro avg	0.99	0.98	0.99	241
weighted avg	0.98	0.98	0.97	241	weighted avg	0.99	0.99	0.99	241
(a)					(b)				
	precision	recall	f1-score	support		precision	recall	f1-score	support
BRCA	0.99	1.00	0.99	95	BRCA	0.99	1.00	0.99	95
COAD	1.00	0.91	0.95	22	COAD	1.00	0.91	0.95	22
KIRC	1.00	1.00	1.00	39	KIRC	1.00	1.00	1.00	39
LUAD	0.95	0.98	0.97	43	LUAD	0.95	0.98	0.97	43
PRAD	1.00	1.00	1.00	42	PRAD	1.00	1.00	1.00	42
accuracy			0.99	241	accuracy			0.99	241
macro avg	0.99	0.98	0.98	241	macro avg	0.99	0.98	0.98	241
weighted avg	0.99	0.99	0.99	241	weighted avg	0.99	0.99	0.99	241
(c)					(d)				
	precision	recall	f1-score	support		precision	recall	f1-score	support
BRCA	0.98	0.99	0.98	95	BRCA	0.99	1.00	0.99	95
COAD	1.00	0.86	0.93	22	COAD	1.00	0.91	0.95	22
KIRC	1.00	1.00	1.00	39	KIRC	1.00	1.00	1.00	39
LUAD	0.91	0.95	0.93	43	LUAD	0.95	0.98	0.97	43
PRAD	1.00	1.00	1.00	42	PRAD	1.00	1.00	1.00	42
accuracy			0.98	241	accuracy			0.99	241
macro avg	0.98	0.96	0.97	241	macro avg	0.99	0.98	0.98	241
weighted avg	0.98	0.98	0.97	241	weighted avg	0.99	0.99	0.99	241
(e)					(f)				
	precision	recall	f1-score	support		precision	recall	f1-score	support
BRCA	1.00	1.00	1.00	95	BRCA	1.00	1.00	1.00	95
COAD	1.00	1.00	1.00	22	COAD	1.00	1.00	1.00	22
KIRC	0.97	1.00	0.99	39	KIRC	0.97	1.00	0.99	39
LUAD	1.00	0.98	0.99	43	LUAD	1.00	0.98	0.99	43
PRAD	1.00	1.00	1.00	42	PRAD	1.00	1.00	1.00	42
accuracy			1.00	241	accuracy			1.00	241
macro avg	0.99	1.00	1.00	241	macro avg	0.99	1.00	1.00	241
weighted avg	1.00	1.00	1.00	241	weighted avg	1.00	1.00	1.00	241
(g)									

Fig. 5. Classification report (a) Logistic regression (b) Naive Bayes (c) Random forest (d) K nearest neighbor (e) Decision tree (f) Support vector machine (g) Proposed approach.

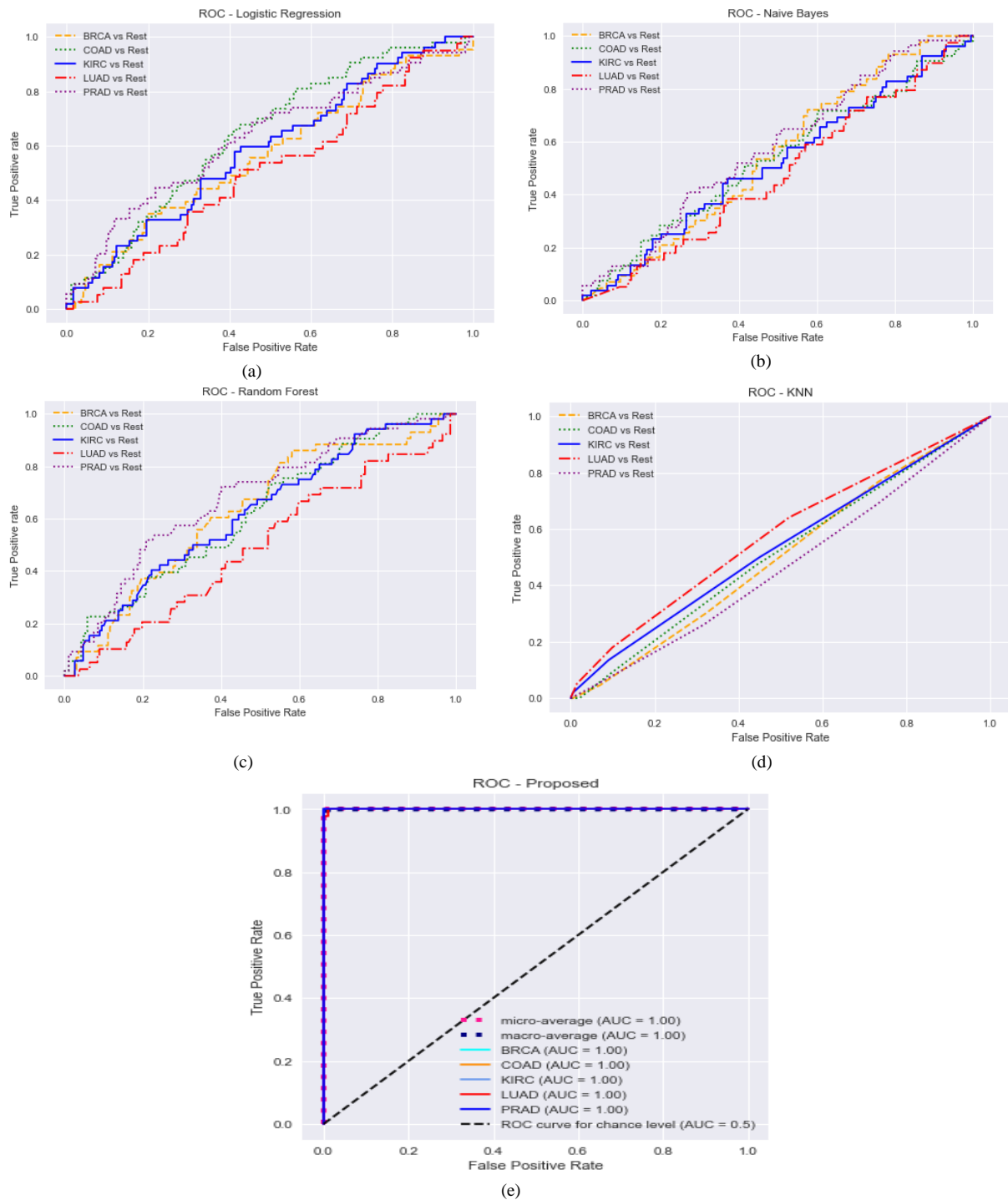


Fig. 6. Receiver operating characteristics curve – One-vs-rest (OvR) (a) Logistic regression (b) Naive Bayes (c) Random forest (d) K nearest neighbor (e) Proposed approach.

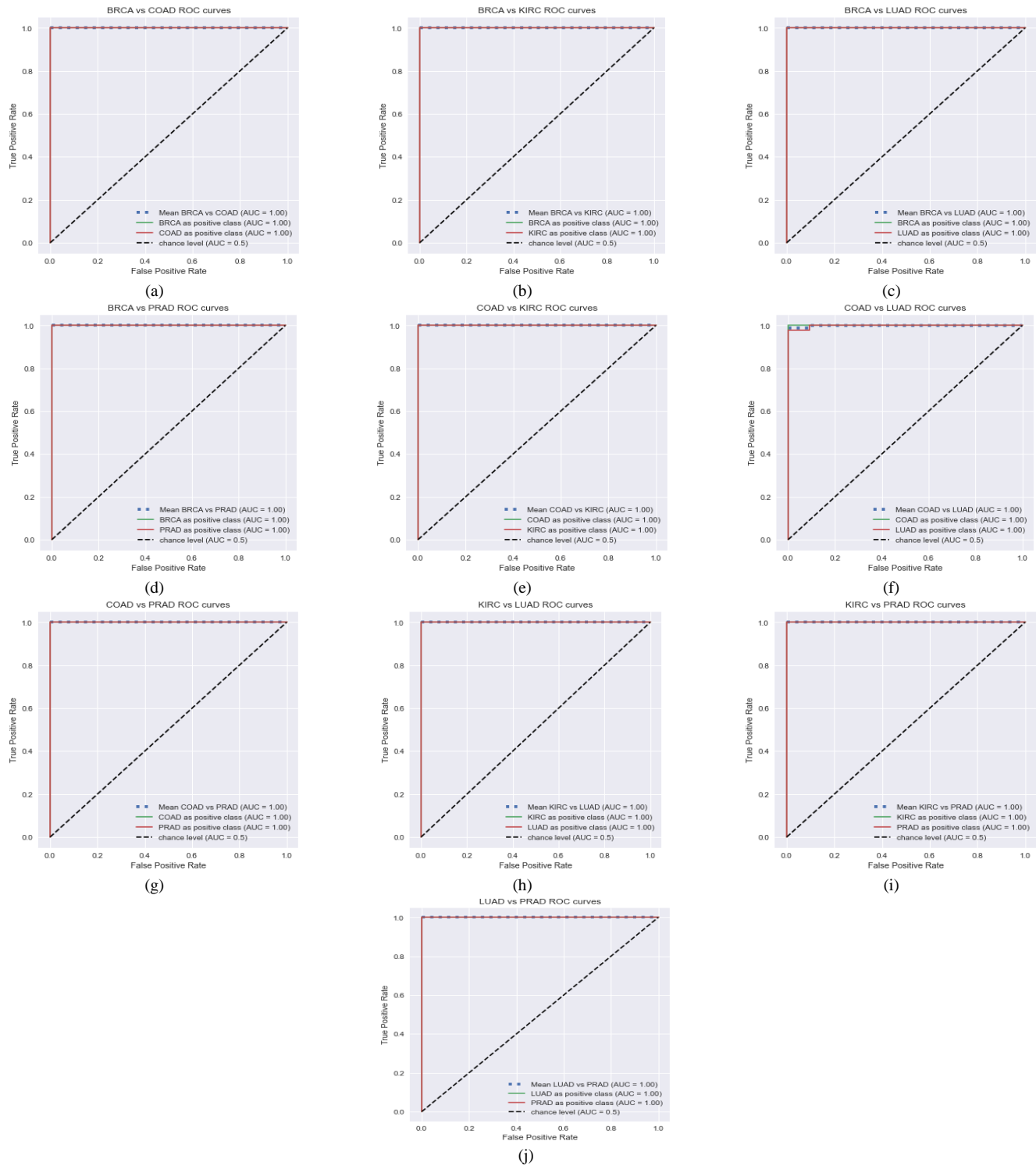


Fig. 7. Receiver Operating Characteristics Curve of Proposed Approach – One-vs-One (OvO) (a)BRCA vs. COAD (b) BRCA vs. KIRC (c) BRCA vs. LUAD (d) BRCA vs. PRAD (e) COAD vs. KIRC (f) COAD vs. LUAD (g) COAD vs. PRAD (h) KIRC vs. LUAD (i) KIRC vs. PRAD (j) LUAD vs. PRAD.

### H. State-of-the-Art Analysis

The state-of-the-art analysis with respect to the reported results of existing cancer RNA classification systems is tabulated in Table III. The proposed approach is compared with optimized deep learning, ensemble classifier, SVM, grouping genetic algorithm, marker gene selection,

dimensionality reduction with neural network, and dimensionality reduction with SVM. It is evident that the proposed approach surpasses the existing cancer classification systems. The reason behind the significant performance is that the curse of dimensionality problem existing in gene sequence data has been overcome using the feature extraction process

and extracted features are utilized for the ensemble classification task. Furthermore, a hard voting classifier has been built using the combination of best-performing classifiers that are chosen based on the trial-and-error process. Thus, the superiority of the proposed approach has been proved.

TABLE III. STATE-OF-THE-ART ANALYSIS

Method	Year	Accuracy (%)
Optimised deep learning [4]	2020	96.9
Ensemble classifier [26]	2020	93.3
Support vector machine [27]	2020	97.37
Grouping genetic algorithm [28]	2020	98.81
Marker gene selection [23]	2021	97.0
PCA-NN [30]	2023	96.6
PCA-SVM [30]	2023	96.5
Proposed	-	<b>99.59</b>

## V. CONCLUSION

The study successfully classified the RNA cancer types from a huge database using the proposed voting ensemble classifier approach. The RNA cancer sequence features were extracted using feature extraction process of PCA to reduce the dimension of the sequence data. The extracted features were used for ensemble classification model building and a hard voting ensemble classifier was effectively applied. In this work a dataset from the UCI Repository was used that includes 801 samples and 20,531 attributes representing five forms of cancer (Breast, Kidney, Colon, Lung, and Prostate). The proposed system used to find an ideal response for the classification of cancer RNA sequences. The accuracy percentage for ensemble categorization is 99.59%. The ROC analysis had been performed with respect to one versus one class and one versus rest of the classes. It is evident that the AUC for the proposed approach is high. Furthermore, the state-of-the-art analysis proved that the proposed ensemble approach outperforms the existing RNA cancer classification systems. In future, the work can be improved by employing a wider variety of exhaustive and thorough techniques, which might be used with other kinds of high-dimensional datasets.

## ACKNOWLEDGMENT

All data were collected and handled in accordance with ethical standards, including anonymization and secure storage, to ensure the protection of participants' privacy and confidentiality.

The dataset used and analyzed during the current study are available in the UCI Machine Learning Repository, <https://archive.ics.uci.edu/ml/datasets/gene+expression+cancer+RNA-Seq>. The data is available publicly and can be used by the machine learning community for the empirical analysis of machine learning algorithms.

Moreover, the first author receives the grant under FDP scheme of UGC India. The second Author has no conflict of Interest.

## REFERENCES

- [1] S. Wesolowski, M. R. Birtwistle, G. A. Rempala, A comparison of methods for RNA-seq differential expression analysis and a new empirical Bayes approach, *Biosensors* 3 (3) (2013) 238–258.
- [2] A. Conesa, P. Madrigal, S. Tarazona, D. Gomez-Cabrero, A. Cervera, A. McPherson, M. W. Szczesniak, D. J. Gaffney, L. L. Elo, X. Zhang, et al., A survey of best practices for RNA-seq data analysis, *Genome Biology* 17 (1) (2016) 1–19.
- [3] D. Goksuluk, G. Zararsiz, S. Korkmaz, V. Eldem, G. E. Zararsiz, E. Ozcetin, A. Ozturk, A. E. Karaagaoglu, Mlseq: Machine learning interface for RNA-sequencing data, *Computer methods and programs in biomedicine* 175 (2019) 223–231.
- [4] N. E. M. Khalifa, M. H. N. Taha, D. E. Ali, A. Slowik, A. E. Hassaniien, Artificial intelligence technique for gene expression by tumor RNA-seq data: a novel optimised deep learning approach, *IEEE Access* 8 (2020) 22874–22883.
- [5] J. Wu, C. Hicks, Breast cancer type classification using machine learning, *Journal of personalised medicine* 11 (2) (2021) 61.
- [6] S. Shamshirband, M. Fathi, A. Dehzangi, A. T. Chronopoulos, H. Alinejad-Rokny, A review on deep learning approaches in healthcare systems: Taxonomies, challenges, and open issues, *Journal of Biomedical Informatics* 113 (2021) 103627.
- [7] D. Sachin et al., Dimensionality reduction and classification through PCA and LDA, *International Journal of Computer Applications* 122 (17) (2015).
- [8] R. Zhang, T. Du, S. Qu, A principal component analysis algorithm based on dimension reduction window, *IEEE Access* 6 (2018) 63737–63747.
- [9] S. Ramroach, A. Joshi, M. John, Optimisation of cancer classification by machine learning generates an enriched list of candidate drug targets and biomarkers, *Molecular omics* 16 (2) (2020) 113–125.
- [10] B. Gunasundari, S. Arun, Ensemble classifier with hybrid feature transformation for high dimensional data in healthcare, in *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, IEEE, 2022, pp. 886–892.
- [11] Y. Xu, Z. Yu, W. Cao, C. P. Chen, A novel classifier ensemble method based on subspace enhancement for high-dimensional data classification, *IEEE Transactions on Knowledge and Data Engineering* 35 (1) (2021) 16–30.
- [12] G. Zararsiz, D. Goksuluk, B. Klaus, S. Korkmaz, V. Eldem, E. Karabulut, A. Ozturk, voomdda: discovery of diagnostic biomarkers and classification of RNA-seq data, *PeerJ* 5 (2017) e3890.
- [13] A. Ishii, K. Yata, M. Aoshima, Geometric classifiers for high-dimensional noisy data, *Journal of Multivariate Analysis* 188 (2022) 104850.
- [14] N. Song, K. Wang, M. Xu, X. Xie, G. Chen, Y. Wang, Design and analysis of ensemble classifier for gene expression data of cancer, *Adv. Genet. Eng* 5 (2015).
- [15] A. McDermaid, X. Chen, Y. Zhang, C. Wang, S. Gu, J. Xie, Q. Ma, A new machine learning-based framework for mapping uncertainty analysis in RNA-seq read alignment and gene expression estimation, *Frontiers in genetics* 9 (2018) 313.
- [16] G. Zararsiz, D. Goksuluk, S. Korkmaz, V. Eldem, G. E. Zararsiz, I. P. Duru, A. Ozturk, A comprehensive simulation study on classification of RNA-Seq data, *PLoS one* 12 (8) (2017) e0182507.
- [17] Y. Guo, S. Liu, Z. Li, X. Shang, Towards the classification of cancer subtypes by using cascade deep forest model in gene expression data, in *2017 IEEE international conference on bioinformatics and biomedicine (BIBM)*, IEEE, 2017, pp. 1664–1669.
- [18] S. Ramroach, M. John, A. Joshi, The efficacy of various machine learning models for multiclass classification of RNA-seq expression data, in *Intelligent Computing: Proceedings of the 2019 Computing Conference, Volume 1*, Springer, 2019, pp. 918–928.
- [19] Y. Xiao, J. Wu, Z. Lin, X. Zhao, A semi-supervised deep learning method based on stacked sparse auto-encoder for cancer prediction using RNA-Seq data, *Computer methods and programs in biomedicine* 166 (2018) 99–105.

- [20] P. Ryvkin, Y. Y. Leung, L. H. Ungar, B. D. Gregory, L.-S. Wang, Using machine learning and high-throughput RNA sequencing to classify the precursors of small non-coding RNAs, *Methods* 67 (1) (2014) 28–35.
- [21] H. R. Hassanzadeh, J. H. Phan, M. D. Wang, A multimodal graph-based semi-supervised pipeline for predicting cancer survival, in 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE, 2016, pp. 184–189.
- [22] A. M. McCarthy, Y. Liu, S. Ehsan, Z. Guan, J. Liang, T. Huang, K. Hughes, A. Semine, D. Kontos, E. Conant, et al., Validation of breast cancer risk models by race/ethnicity, family history and molecular subtypes, *Cancers* 14 (1) (2021) 45.
- [23] B. Aevermann, Y. Zhang, M. Novotny, M. Keshk, T. Bakken, J. Miller, R. Hodge, B. Lelieveldt, E. Lein, R. H. Scheuermann, A machine learning method for the discovery of minimum marker gene combinations for cell type identification from single-cell RNA sequencing, *Genome Research* 31 (10) (2021) 1767–1780.
- [24] R. Zhu, Z. Wang, N. Sogi, K. Fukui, J.-H. Xue, A novel separating hyperplane classification framework to unify nearest-class-model methods for high-dimensional data, *IEEE transactions on neural networks and learning systems* 31 (10) (2019) 3866–3876.
- [25] B. Pes, Ensemble feature selection for high-dimensional data: a stability analysis across multiple domains, *Neural Computing and Applications* 32 (10) (2020) 5951–5973.
- [26] M. O. Arowolo, M. Adebisi, A. Adebisi, O. Okesola, Pca model for RNA-seq malaria vector data classification using KNN and decision tree algorithm, in 2020 international conference in mathematics, computer engineering and computer science (ICMCECS), IEEE, 2020, pp. 1–8.
- [27] Z. Yu, Z. Wang, X. Yu, Z. Zhang, et al., Rna-seq-based breast cancer subtypes classification using machine learning approaches, *Computational intelligence and neuroscience* 2020 (2020).
- [28] P. Garcia-Diaz, I. S´anchez-Berriel, J. A. Mart´inez-Rojas, A. M. Diez-Pascual, Unsupervised feature selection algorithm for multiclass cancer classification of gene expression RNA-seq data, *Genomics* 112 (2) (2020) 1916–1925.
- [29] M. A. Mohammed, A. Lakhan, K. H. Abdulkareem, B. Garcia-Zapirain, A hybrid cancer prediction based on multi-omics data and reinforcement learning state action reward state action (sarsa), *Computers in Biology and Medicine* 154 (2023) 106617.
- [30] M. O. Arowolo, M. O. Adebisi, A. A. Adebisi, A genetic algorithm approach for predicting ribonucleic acid sequencing data classification using knn and decision tree, *TELKOMNIKA (Telecommunication Computing Electronics and Control)* 19 (1) (2021) 310–316.
- [31] M. O. Arowolo, M. Adebisi, A. A. Adebisi, J. OKesola, Predicting RNA-seq data using genetic algorithm and ensemble classification algorithms, *Indonesian Journal of Electrical Engineering and Computer Science* 21 (2) (2021) 1073–1081.
- [32] M. Ramamurthy, I. Krishnamurthi, S. Vimal, Y. H. Robinson, Deep learning-based genome analysis and NGS-RNA II identification with a novel hybrid model, *Biosystems* 197 (2020) 104211.
- [33] M. Mohammed, H. Mwambi, I. B. Mboya, M. K. Elbashir, B. Omolo, A stacking ensemble deep learning approach to cancer type classification based on tcga data, *Scientific reports* 11 (1) (2021) 1–22.
- [34] M. O. Arowolo, M. Adebisi, A. A. Adebisi, An efficient PCA ensemble learning approach for prediction of RNA-seq malaria vector gene expression data classification, *International Journal of Engineering Research and Technology* 13 (1) (2020) 163–169.
- [35] M. F. Kabir, T. Chen, S. A. Ludwig, A performance analysis of dimensionality reduction algorithms in machine learning models for cancer prediction, *Healthcare Analytics* 3 (2023) 100125.
- [36] K. Pradhan, P. Chawla, Medical internet of things using machine learning algorithms for lung cancer detection, *Journal of Management Analytics* 7 (4) (2020) 591–623.
- [37] A. A. Osuwa, H. Oztoprak, Importance of continuous improvement of machine learning algorithms from a health care management and management information systems perspective, in 2021 International Conference on Engineering and Emerging Technologies (ICEET), IEEE, 2021, pp. 1–5.
- [38] F. Alharbi, A. Vakanski, Machine learning methods for cancer classification using gene expression data: A review, *Bioengineering* 10 (2) (2023) 173.
- [39] W. M. Ead, M. A. Abdelazim, M. M. Nasr, Feedforward deep learning optimiser-based RNA-Seq women’s cancers detection with a hybrid classification models for biomarker discovery, *International Journal of Advanced Computer Science and Applications* 13 (12) (2022).
- [40] M. A. Talukder, M. M. Islam, M. A. Uddin, A. Akhter, K. F. Hasan, M. A. Moni, Machine learning-based lung and colon cancer detection using deep feature extraction and ensemble learning, *Expert Systems with Applications* 205 (2022) 117695.
- [41] K. Ferles, Y. Papanikolaou, ‘cancer types: RNA sequencing values from tumour samples/tissues, Distributed by Mendeley (2018).
- [42] Japkowicz, N., Shah, M. (2015). Performance Evaluation in Machine Learning. In: El Naqa, I., Li, R., Murphy, M. (eds) Machine Learning in Radiation Oncology. Springer, Cham.
- [43] Visa, Sofia & Ramsay, Brian & Ralescu, Anca & Knaap, Esther. (2011). Confusion Matrix-based Feature Selection. *CEUR Workshop Proceedings*. 710. 120-127.
- [44] M. H. Zweig, G. Campbell, Receiver-operating characteristic (roc) plots: a fundamental evaluation tool in clinical medicine, *Clinical chemistry* 39 (4) (1993) 561–577.
- [45] Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve, *Radiology*. 1982;143(1):29–36.
- [46] Student S, Fajarewicz K. Stable feature selection and classification algorithms for multiclass microarray data. *Biol Direct*. 2012 Oct 2;7:33. doi: 10.1186/1745-6150-7-33. PMID: 23031190; PMCID: PMC3599581.



# Historical Building 3D Reconstruction for a Virtual Reality-based Documentation

Ahmad Zainul Fanani, Arry Maulana Syarif

Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang, Indonesia

**Abstract**—An innovative preservation approach was proposed to document historical buildings in 3D model, and to present it virtually. The approach was applied to the Lawang Sewu building, one of the architectural masterpieces that is part of Indonesian history. Virtual Reality (VR) technology was used to create a Lawang Sewu VR application program that allows users to virtually walk around the building. A new method for 3D reconstruction was proposed, where data of photo, video and miniature documentation, as well as notes collected from observations were used as the main reference. Meanwhile, architectural record data was used in cases where information cannot be obtained through the main reference. The proposed method focuses on traditional techniques, both at the data acquisition and 3D modelling stages. Poly modelling techniques were chosen for 3D reconstruction. The poly modelling technique was chosen based on its ease and flexibility in controlling the number of polys in 3D models, and was suitable to be applied for repetitive spatial typologies, such as the Lawang Sewu building. After given textures, the 3D model was sent to the VR editor. In addition of running on the desktop platform, Head Mounted Device (HMD) that supports the creation of an immersive experience, was also chosen to run the Lawang Sewu VR. The evaluation carried out to measure the level of similarity of the 3D model to the original building and the sensation of an immersive experience felt by the user shows good achievements.

**Keywords**—Virtual reality; immersive presentation; 3D reconstruction; historical heritage building preservation

## I. INTRODUCTION

Lawang Sewu is a historic building that became one of the markers of the city of Semarang, Central Java, Indonesia. This building is the work of the famous Dutch architect, C. Citroen from the J.F. Firm. Klinkhamer and B.J. Quendag in 1903 and completed in 1907 for the headquarters of the Dutch colonial railway company, or Nederlandsch Indische Spoorweg Naatschappij. The Lawang Sewu building was designed to have a lot of windows and doors as an air circulation system, and this is where the term Lawang Sewu, which in Javanese means a thousand doors, came from. The Lawang Sewu complex which stands on an area of about 18,232 square meters consists of five buildings, which are buildings A, B, C, D, E, and lavatory. Fig. 1 shows the photo collection of the Lawang Sewu building.

An innovative approach was proposed to support the preservation of the Lawang Sewu building as an architectural masterpiece which is currently functioning as a tourist destination. The proposed approach allows users who have not had the opportunity to visit the Lawang Sewu building can see the architectural details of the building virtually. VR

technology based on 3D reconstruction was proposed to create a Virtual Lawang Sewu program application that can document architectural details of buildings in 3D format. Therefore, users can walk around the building virtually. VR is an immersive technology that allows users to interact subjectively with the virtual world so that they can feel the sensation of their physical present. VR is an environment that is displayed in the form of media that is able to create a sensation for users who seem to be physically in their surroundings [1], and 3D reconstruction techniques are developing rapidly to meet the needs of geometric 3D models for the film, game and virtual environment industries, such as works [2-5]. In this study, for the purpose of the documentation and virtual presentation, the Lawang Sewu building was reconstructed into a 3D model to be applied to VR applications that can be run in the desktop and HMD platforms. The main problem was to reconstruct the building into a 3D model as precisely as possible. In order to maintain the number of polys in the 3D model, traditional 3D modelling techniques was chosen to create the 3D model. The technique has consequences for the process of selecting data sources and analyzing them, which are also carried out manually, such as carrying out careful photography sessions to obtain information from the building profile, or measuring every detail of the building profile directly or other approaches. After that, carry out an analysis of the information obtained to calculate the shape and size of the building. The proposed traditional 3D modeling technique can be used to reconstruct buildings into 3D models with precision.

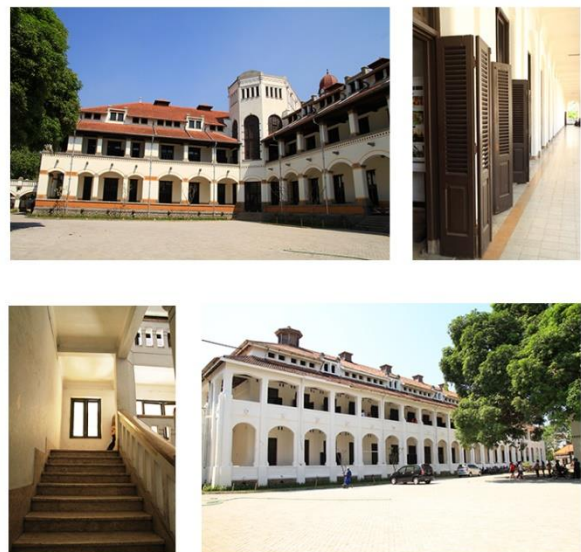


Fig. 1. The Lawang Sewu building.

The remainder of this paper is organized as follows: Section II discusses a review of some related work, Section III describes the proposed methodology which includes techniques in the data collection, techniques in the 3D reconstruction and 3D model texturing, VR programming, and evaluation. Section IV analyzes results. Section V covers the conclusion.

## II. RELATED WORK

There are various types of sensations of user presence in a virtual world (immersion), which are tactical to feel experience in carrying out tactical operations that require skill, strategic to feel mental challenges, narrative to feel being in a story, spatial to feel being in the real world, psychological to feel anxiety over the game with real life, and sensory to feel being in a unity of time and place based on the virtual environment [6]. VR consists of hardware components including computer sets, sensor embodiments (head mounted displays, binocular orientation monitors, and monitors), process acceleration cards, tracking systems, input devices, and software components including 3D modelling software, graphics, audio, and virtual reality simulation [7].

3D reconstruction is one of the challenges in developing VR applications. The challenge in 3D reconstruction is to formulate the right method in creating a 3D model that is as close as possible to the original object [8]. 3D reconstruction research was conducted for building objects [9-10], underwater environments [1, 5], small objects [11], and other objects. 3D reconstruction can be grouped by time (time-based reconstruction). For example, 3D reconstruction that aims to visualize objects based on their current construction such as works [2-3, 12-15], and 3D reconstructions that aim to visualize objects that have been damaged into their intact form such as works [16]. Research on 3D reconstruction for the Coliseum building in Rome, Italy, was carried out by [17], and the Great Wall in China by [18]. Both studies used tourism photo data from the [www.flickr.com](http://www.flickr.com) site. The Coliseum building model was generated from 2106 photos, while the Great Wall model was generated from 120 photos. Further, it was explained that the challenge of this research is matching and 3D reconstruction of information from hundreds or thousands of photos consisting of variations in perspective, illumination, weather, resolution, and others that have the potential for clutter and outliers. The real-time room environment reconstruction technique uses an octre-based surface representation for Kinect Fusion, where the space is represented as a signed distance function and stored as a uniform grid of voxels [18].

Technological developments have enabled VR application programs to be presented through stationary displays (desktop-VR or CAVE), head-based displays (HMD-VR or smartphone-VR), and Hand-based displays (Handheld VR) [9]. This supports the so-called Second Chance Tourism that utilizes digital technology, such as VR technology, which allows tourists to get the experience of visiting tourist sites without physically having to be on site [4]. On the other hand, the appearance of an attractive 3D model is one aspect of building the absorptive experience that the user gets, and the absorptive experience has an influence on the level of immersive felt by the user in a virtual environment [19]. Therefore, the use of VR

technology for the preservation of cultural heritage needs to consider the design of attractive 3D models. Meanwhile, multi-experiential which includes learning and educational experiences, including emotional experiences, has become part of the existence of cultural heritage preservation [20]. Therefore, VR application programs also need to be designed to be able to provide various experiences for users while in a virtual environment.

## III. METHODOLOGY

Workflow used in this study was designed based on eight challenges in developing the model of tangible 3D-based cultural heritage preservation identified by [8], which are time-based 3D reconstruction, typology, 3D reconstruction method, application category, research objective, data management, presentation method and research evaluation. Based on the time, there are two types of time in the 3D based preservation of historical objects, which are 3D reconstruction based on the current environment that uses data from the current condition of historical objects, and based on the past environment that uses data from historical objects that have been damaged, even extinct [8]. The proposed 3D reconstruction for the Lawang Sewu building referred to the current physical condition of the building, where the current physical data of the building was analyzed for use in its 3D modelling. Typological analysis was carried out to design data collection techniques based on the detailed characteristics of the building. Data was collected through: 1) direct observation by documenting the architectural details of the Lawang Sewu building in records, photos and videos; 2) the building miniature observation; 3) the building blueprint analysis. After determining the time-based 3D reconstruction method and performing typological analysis, the data management stage was carried out to design the data flow for the process before, after and during the 3D reconstruction. Based on the chosen time-based 3D Reconstruction, typology analysis, and data management design, traditional 3D modelling techniques was chosen to create the Lawang Sewu 3D model. The technique was chosen based on its ease and flexibility in controlling the number of polys in 3D modelling to reconstruct buildings with repetitive spatial typologies, such as the Lawang Sewu building.

In addition to measuring the physical building, a new method proposed to measure the area of the building in 3D reconstruction is to use the size of one of small elements of the building, and then count the number of the chosen element in a room. For example, given a tile chosen for the element to measure area of a room with a size of 30 x 30 cm; the length and width of the room uses 10 tiles; and the area of the room is 900 x 900 cm. The tile order index is also used to identify the position of other elements, such as doors, windows, poles. For example, given the door in a room, the door is in the order of tiles 4 to 7, so the door size is around 120 cm. The proposed method uses calculations based on the size of the element of the building that is the reference for measurement and its number. Moreover, the proposed method can facilitate texturing work, where the texturing process for building elements, such as tiles, can be appropriate based on the number of elements. On the other hand, there are building elements that still require size data based on architectural records, such as building height, or building elements that require physical

measurements for comparison data or because their position is disconnected from the elements used as measurement references, such as poles in the building yard.

Accurate building size is not the target, because this 3D reconstruction project aims to document historical buildings for virtual presentation purposes. Therefore, visual similarity is the target, and not the accuracy of the building's size. Later, evaluation results show that this method is effective for 3D reconstruction which can create 3D objects with a size scale that is close to real objects. The workflow of the Lawang Sewu VR development consists of five stages, which are data collection, 3D reconstruction, 3D texturing, VR programming, and evaluation.

### A. Data Collection

In addition to direct observation to the location of the Lawang Sewu building, the data sources used include a collection of photos, videos, miniatures and blueprints of the building. First of all, direct observation to the location was conducted in order to get a general picture of the environment, such as the layout, shape and structure of the building. Furthermore, an analysis of the building blueprint was carried out to sharpen the understanding of the building information from an architectural perspective. The building blueprint is displayed in one of the rooms used as a museum in the Lawang Sewu building. The building miniature which is also displayed in the museum was a medium to better understand the layout, shape, and structure of the building in a 3D perspective. Furthermore, photo and video sessions were conducted on every detail of the building, including physical measurements of the building. Physical measurements of buildings were not carried out on all building constructions, but on certain parts, such as doors, windows, stairs, tiles, and several other parts. Fig. 2 shows some blueprints of the Lawang Sewu building, while Fig. 3 shows the illustration of the miniature buildings.

All data sources were treated like puzzles in working on 3D reconstruction, where data serve to complement each other. The data acquisition method in 3D reconstruction proposed in this study adopts a puzzle game.

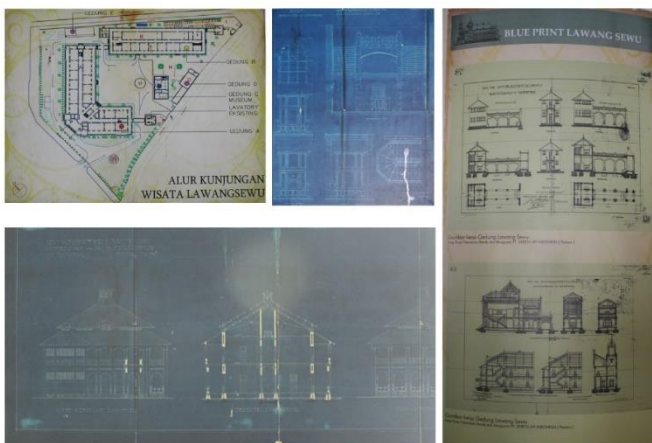


Fig. 2. Example of the blueprint of the Lawang Sewu building



Fig. 3. Example of the blueprint of the Lawang Sewu building

### B. 3D Reconstruction

The Lawang Sewu building consists of buildings A, B, C, D, and E, including the basement which is located under building B. In this study, 3D reconstruction was targeted at the two main buildings, which are building A and B, including the Lavatory connected by a bridge to building A. The 3D reconstruction phase started from building B, continued with building A and Lavatory including the connecting bridge between them. Fig. 4 shows the layout of the Lawang Sewu building based on its blue print.



Fig. 4. The blueprint of the buildings.

Building B has a size of 22x77 m<sup>2</sup> or an area of 4,145.21 m<sup>2</sup> with two main floors and one roof space. 3D reconstruction begun by identifying the size of the one room at the very end. The design of building B has a repeating pattern of rooms with almost all rooms being the same size. There are different sized rooms that are twice the size of the other rooms. The 3D reconstruction process in building B which consists of three floors was started from the first floor by identifying the size of

the tiles and counting their number like cells in a matrix, where the number of cells in a row represents the width of the room, and the number of cells in a column represents the length of the room. Furthermore, as long as it is still accessible by hand, physical measurements of building elements were also carried out, such as the thickness of walls, the size of doors, windows, poles and stairs. In order to reduce the amount of poly, the ceiling of the room was not 3D reconstructed, but created using image textures from the original photo. Meanwhile, the height of the room was identified using architectural records data. After one room at the very end has been reconstructed in 3D, the process continues with the next room, and so on until the room at the other end. The method was also applied to obtain size, shape and layout data of other objects in the building. Fig. 5 shows illustration of using the size and number of tiles to identify the area of the room and the position of elements in it, while Fig. 6 shows example of an object that require physical measurement.

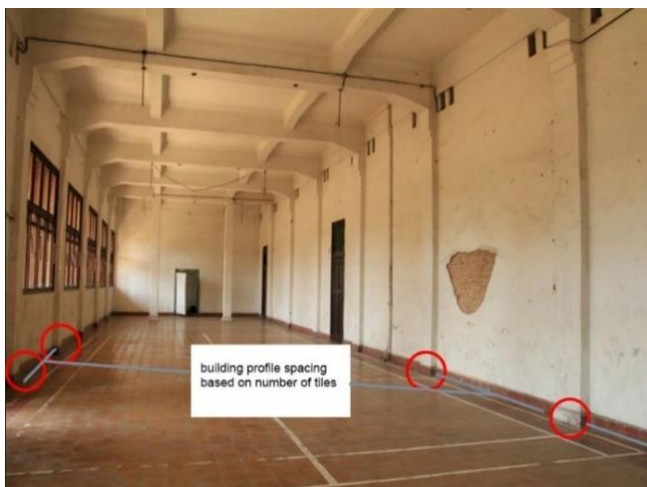


Fig. 5. Illustration of using the size and number of tiles to determine the area of the room and the position of the profiles in it.



Fig. 6. Example of an object that require physical measurement.

After obtaining the data and information of building B, the process continued with the creation of a 3D model of the building using the 3Ds Max application program. Stages in the process of making 3D models were carried out as in the stage of data acquisition for buildings. Starting from the very end of the room, then it was duplicated to complete the design of the

first floor of building B, and continued with corridors and terraces. Other editing, such as merging two rooms into one room, and making stairs objects were carried out with the support of photo documentation data. After the first floor was completed, the 3D reconstruction was continued to the second and third floors using the same methods and techniques. Meanwhile the 3D reconstruction for the roof of the building was carried out using data obtained based on observations and analysis on the miniature building. Next was applying a cleaning process to remove unnecessary vertices and polygons.



Fig. 7. The comparison illustration between the original photo of building B and the 3D model of building B.

The cleaning process was performed to maintain the number of vertices and polys in the 3D model for its size does not swell and the application program can be lighter when being played. The cleaning process produced a 3D model of building B with a total of 75,209 polygons and a total of 107,619 vertices. Fig. 7 shows a comparison illustration between the original photo of building B and the 3D model of building B. Further, the 3D model of building B was used as a reference in the 3D reconstruction of building A.

Building B has a size of 22x77 m<sup>2</sup> or an area of 4,145.21 m<sup>2</sup> with two main floors and one roof space. Building A is a three-story main building that has a shape like the letter L with an area of 5,473.28 m<sup>2</sup>, and the floor of the office and lobby is tiled with a size of 16x16 m<sup>2</sup>. Meanwhile, the 2-story Lavatory has an area of 242.60 m<sup>2</sup>. The same methods and techniques were applied in the 3D reconstruction of building A. The 3D reconstruction of the bridge connecting building A and Lavatory was carried out by continuing, or embedding, in the 3D model of building A. The wall of building B which is at the end of the building that connects to building A was used as the starting point.

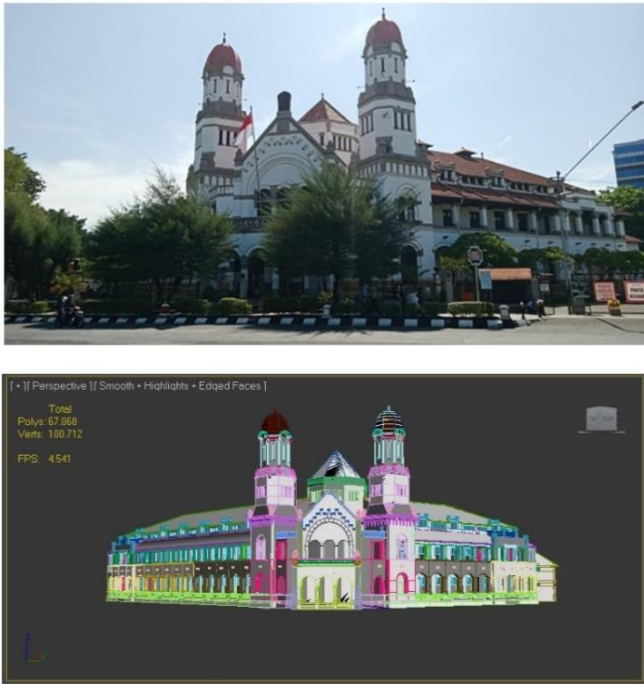


Fig. 8. Comparison illustration between the original photo of building A and the 3D model of building A.

The cleaning process produced a 3D model of building A and the lavatory with a total of 67.068 polys and a total of 100.712 vertices. Fig. 8 shows a comparison illustration between the original photo of building B and the 3D model of building B. The 3D model of the Lawang Sewu building consisting of buildings A, B, and Lavatory produced has a poly count of 142,227 and a vertex of 208,331. Fig. 9 shows an illustration of the results of making the model.

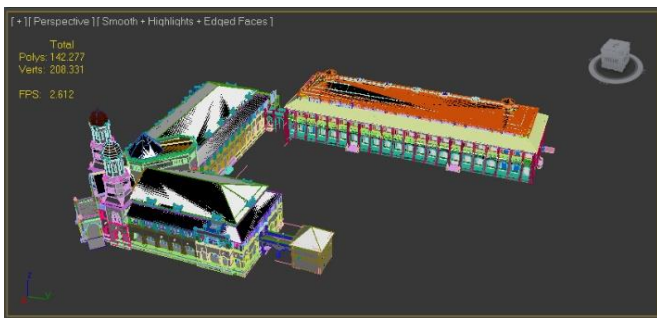


Fig. 9. A Results of the 3D model of the Lawang Sewu building consisting of buildings A, B, and Lavatory.

### C. Texturing

The 3D texturing process was carried out using the Unwrap UVW modifier tool in 3Ds Max. The tool is for applying and controlling more than one texture on various parts of the object. First of all, the details of the surface of the real object were photographed. Furthermore, the 3D object is applied with the unwrap technique to produce a pattern of parts of the object in the form of an outline. The pattern image is saved in PNG format and sent to the Adobe Photoshop application program to be textured using a warp technique based on the photo details

of the real object. Fig. 10 shows an illustration of an unwrap image resulted from the 3Ds Max application program which was then applied to the warp technique in the Adobe Photoshop application program using photo details of the real object.

The 3D texturing process was also carried out using the image texture mapping technique. This technique can keep the number of polygons from swelling in making 3D models more realistic and attractive, in which 3D objects are applied with color patterns [21]. Some objects, such as doors and windows, were manipulated using the image texture bump mapping technique. 3D models for doors and windows were formed using boxes, a primitive shape type, consisting of six polygons and eight vertices. Details of the original object profile were photographed, then edited using an image editor application program, Adobe Photoshop. Image of the object profile, then used to give texture to the object. Fig. 11 shows an illustration of the application of the image texture mapping technique to create a 3D door model. Details of the door profile, including the surface look realistic although it is built from a 3D box object consisting of a small number of polygons and vertices with a flat surface.



Fig. 10. Illustration of an unwrap image resulted from the 3Ds Max application program which was then applied to the warp technique in the Adobe Photoshop application program (a) using photo details of the real object (b).



Fig. 11. Illustrations of (a) that is the door profile image used in the image texture mapping implementation for a 3D box object (b).

#### D. VR Programming

The Lawang Sewu VR computer program was developed to run on the desktop and the HMD platforms with a consideration that desktop applications are still popular and widely used by users, while HMD-VR applications, although currently gaining popularity, not many users have the devices. A teleportation feature that allows users to change locations from building A to building B, or vice versa, including building floor selection is added to the Lawang Sewu VR. This feature makes it easy for users to get around the building virtually. The challenge in implementing 3D assets into the virtual environment presented through HMD is determining the proportion of users and 3D objects. The traditional technique was used by comparing the proportions of humans and objects in the real environment with the proportions of avatars and objects in the virtual environment through certain poses.

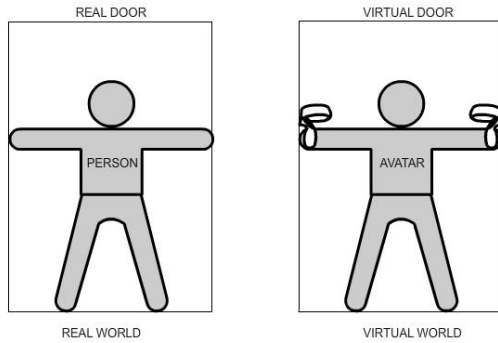


Fig. 12. Illustration of setting a 3D model scale to make it proportional to the original object.

In this case, a person was asked to pose with his arms stretched out in a door of the Lawang Sewu building in the real environment, and the pose was documented through photographs. After that, the person was asked to play Lawang Sewu VR using Oculus Quest 2, and pose in the same location as in the real world. Comparison of photos in the real environment and visualization of user games through casting were the benchmarks in scaling the 3D model so that it is proportional. Fig. 12 shows an illustration of setting a 3D model scale to make it proportional to the original object. Fig. 13 shows user's activities in playing the program using Oculus Quest 2.0, while Fig. 14 shows screenshots of the Lawang Sewu VR played in the desktop platform.



Fig. 13. Illustration of user's activities in playing the Lawang Sewu VR using Oculus Quest 2 HMD.



Fig. 14. Screenshots of the Lawang Sewu VR for desktop.

#### E. Evaluation

User acceptance test was carried out to measure the achievement of the goal of developing the Lawang Sewu VR computer program, which are documenting historical buildings and presenting them virtually using VR technology. The performance of the Lawang Sewu run in desktop-VR and HMD-VR platforms was measured by evaluating the level of visual similarity and the level of area proportionality between the 3D model of buildings and the original objects. An additional evaluation was carried out on the Lawang Sewu HMD-VR which was the level of immersion evaluation in order to measure the sensation of the user presence in the virtual environment of the Lawang Sewu building.

Respondents who work as tour guides were selected based on their profession in taking tourists every day to tour the Lawang Sewu building. They were assumed to have knowledge of the physics of the Lawang Sewu building. The other 15 respondents were those who had visited the Lawang Sewu building in the past month. Out of 30 respondents, 12 of them were women, and the rest were men. The youngest respondent was 11 years old and the oldest was 51 years old. Respondents were asked to rate the performance of the Lawang Sewu VR by providing opinions.

Results of the user acceptance test for each the Lawang Sewu run in desktop-VR and HMD-VR platforms were measured using the Mean Opinion Score (MOS) technique with the following formula, where R is the individual rating for the stimulus given by subject N:

$$MOS = \frac{\sum_{n=1}^N R_n}{N} \quad (1)$$

Results of the calculation of the MOS value were converted to a range of values 0 - 1 representing bad performance, 1.1 - 2.4 representing poor performance, 2.5-3.4 representing good

performance, and 3.5-4 representing excellent performance. The following are statements (S) judged for gaining opinions of users:

- S1: The 3D model of the Lawang Sewu building has a high level of visual similarity to the original building.
- S2: The 3D model of the Lawang Sewu building has a high degree of similarity in size proportion to the original building.
- S3: The sensation of being in the virtual environment of the Lawang Sewu building really feels like being in the original environment.

The user acceptance test was carried out in two sessions, where the first and second sessions were to provide opinions on the Lawang Sewu desktop-VR, and the Lawang Sewu HMD-VR, respectively. In the first session, each respondent was guided in playing the Lawang Sewu desktop-VR. After getting used to playing the application, each respondent was asked to play it by walking along the usual route for an

unlimited duration. After the respondent completes the usual route taken virtually, respondents were asked to provide an opinion on the S1 and S2 statements in the range of values of 1-4 which represent Strongly Disagree, Disagree, Agree and Strongly Agree, on each statement based on their experience when playing the application. Meanwhile, the second session took place the next day. The evaluation mechanism was the same as in the first session, but in this session the respondents used the HMD device. In this session, respondents were asked to provide an opinion on the S1, S2, and S3 statements.

Evaluation in the first session that measured the Lawang Sewu desktop-VR performance resulted that out of 30 respondents, MOS scores for the statement S1 and S2 were 3.5, and 3.3. Meanwhile, evaluation in the second session that measured the Lawang Sewu HMD-VR performance resulted that out of 30 respondents, MOS scores for the statement S1, S2, and S3 were 3.5, 3.2, and 3.6, respectively. Table I and Table II show MOS results of the performance of the Lawang Sewu run in desktop-VR and HMD-VR, respectively.

TABLE I. MOS RESULTS OF THE LAWANG SEWU DESKTOP-VR

Statements	Opinion Score				MOS Score	
	1	2	3	4		
Visual Similarity (S1)	0	0	14	16	3.5	Good
Size Proportionality (S2)	0	0	22	8	3.3	Good

TABLE II. MOS RESULTS OF THE LAWANG SEWU HMD-VR

Statements	Opinion Score				MOS Score	
	1	2	3	4		
Visual Similarity (S1)	0	1	14	15	3.5	Excellent
Size Proportionality (S2)	2	3	13	12	3.2	Good
Immersion Level (S3)	0	1	11	18	3.6	Excellent

#### IV. RESULTS AND DISCUSSION

An application program based on VR technology was developed to document historical buildings and present them virtually. The application program called The Lawang Sewu VR documents the historic Lawang Sewu building located in Indonesia by reconstructing the building into a 3D model and providing a texture similar to the current condition of the building. Furthermore, the Lawang Sewu 3D Model was sent to the game engine editor to be developed into a VR-based application program that allows users to go around the Lawang Sewu building environment virtually. Desktop and HMD platforms were the targets for running The Lawang Sewu VR, considering that many users already have devices to play desktop-based applications, while HMD-based applications provide a strong immersive sensation.

Based on the typology of buildings that have repetitive patterns, data acquisition and 3D reconstruction were more focused on the use of traditional methods and poly modeling techniques. Photo and video data including architectural records obtained through direct observation were used as references for 3D reconstruction. Some of the data that cannot be obtained through observation were collected through the blue print of the building. The solution in the use of the poly modelling technique was proven to be able to control the number of polys and vertices in the details of the curve of the

building in the 3D model of the Lawang Sewu building which includes buildings A, B and Lavatory. The development of the Lawang Sewu VR application program lasted three months which was divided into one month for data acquisition, one and a half months for 3D reconstruction, and half a month for VR programming. The 3D reconstruction phase involves the most human resources. Ten students from Universitas Dian Nuswantoro, were involved in this stage.

The achievement of the goal of documenting the Lawang Sewu building and presenting it virtually was measured based on the visual similarity and size proportionality of the 3D model to the real building. The traditional 3D modelling technique was conducted based on information obtained from carrying out careful photography sessions, measuring every detail of the building profile directly and the building blueprint analysis. The proposed 3D reconstruction technique is appropriate for building objects with profile and visual characteristics that have symmetry patterns that can be easily identified and measured, such as having tiles of the same size, the same distance between building pillars, or others. The proposed 3D modeling technique is proven to be able to appropriately reconstruct buildings with repetitive spatial typologies, such as the Lawang Sewu building. The user acceptance test measured using the MOS technique on the Lawang Sewu desktop-VR shows that both visual similarity and size proportionality reach a good level. Meanwhile, the

visual similarity in the Lawang Sewu HMD-VR reached an excellent level, and size proportionality reached a good level. This achievement shows that the documentation of the Lawang Sewu building into a 3D model format with visuals and sizes that are close to real objects can be carried out well, and VR technology that supports users around the location virtually can perform well. Especially in the Lawang Sewu HMD-VR, the measurement of the immersive level, the sensation of the user's presence in the virtual world, can achieve an excellent MOS score. Although it still requires further testing, it can be assumed that in 3D reconstruction of historic buildings as assets in HMD-based VR application programs, the relationship between visual similarity and size proportionality has a significant role on the level of immersive perceived by the user. Through some light discussions after trying the Lawang Sewu HMD-VR, some users tried to compare the visuals and sizes of several building elements in a 3D model with real objects, such as doors, windows, stairs and others.

## V. CONCLUSION AND FUTURE WORK

The 3D reconstruction method and the use of VR technology proposed in this study are proven to be able to document historic buildings in 3D model format and present them virtually and interactively. However, the 3D reconstruction technique used in this research is appropriate for building objects with profile and visual characteristics that have repetitive spatial typologies, such as the Lawang Sewu building.

At this time, the functionality of the Lawang Sewu VR is still limited to documentation and virtual presentation of the physical building based on its current condition. The story telling functionality that is able to visualize the physical and historical conditions in the past is the target for further development, including the addition of a multi-user feature that allows more than one user to interact in a virtual Lawang Sewu environment.

## ACKNOWLEDGMENT

Thank to Ministry of Research, Technology, and Higher Education of The Republic Indonesia for financial support through the second year of Applied Research Grant 2022 (Hibah Penelitian Terapan 2022).

## REFERENCES

- [1] P.N. Andono, I.K.E. Purnama, M. Hariadi, T. Watanabe, and K. Kondo, "3D surfaces reconstruction of seafloor images using multiview camera based on image registration". 2012 International Conference on Multimedia Computing and Systems, pp. 803-808, 2013. Doi: 10.1109/ICMCS.2012.6320131.
- [2] L. Argyriou, E. Economou, and V. Bouki, "360-degree interactive video application for cultural heritage education". 3rd Annual International Conference of the Immersive Learning Research Network. Pp. 297-304, 2017. Doi: 10.3217/978-3-85125-530-0-44.
- [3] A. Bachvarov, D. Chotrov, Y. Yordanov, and Z. Uzunova, "Conceptual model of the VR module for virtual plaza for interactive presentation of Bulgarian cultural heritage", AIP Conference Proceedings 2172, pp. 1-5, 2019. Doi: 10.1063/1.5133585.
- [4] A. Bec, B. Moyle, V. Schaffer, and K. Timms, "Virtual reality and mixed reality for second chance tourism", Tourism Management, vol. 83, 104256, 2021. Doi: 10.1016/j.tourman.2020.104256.
- [5] G. Bianco, A. Gallo, F. Bruno, and M. Muzzupappa, "A comparative analysis between active and passive techniques for underwater 3d reconstruction of close-range objects", Sensors, vol. 13, no. 8, pp. 11007-11031, 2013. Doi: 10.3390/s130811007.
- [6] S. Mandal. "Brief introduction of virtual reality & its challenges", International Journal of Scientific & Engineering Research, vol. 4, no. 4, pp. 304-309, 2013.
- [7] M.O. Onyesolu, and F.U. Eze, "Understanding virtual reality technology: Advances and applications." In (Ed.), Advances in Computer Science and Engineering. IntechOpen, 2011. Doi: 10.5772/15529.
- [8] A.Z. Fanani, K. Hastuti, A.M. Syarif, and P.W. Harsanto, "Challenges in developing virtual reality, augmented reality and mixed-reality applications: Case studies on a 3d-based tangible cultural heritage conservation", International Journal of Advanced Computer Science and Applications, vol. 12, no. 11, pp. 219-227, 2021. Doi: 10.14569/IJACSA.2021.0121126.
- [9] Y. Zhang, H. Liu, S-C. Kang, and M. Al-Hussein, "Virtual reality applications for the built environment: Research trends and opportunities. Automation in Construction", 118, 103311, 2020. Doi: 10.1016/j.autcon.2020.103311.
- [10] M. Zeng, F. Zhao, J. Zheng, and X. Liu, "Octree-based fusion for realtime 3D reconstruction. Graphical Models", vol. 75, no. 3, 2013.
- [11] V. Evgenikou, and A. Georgopoulos, "Investigating 3D reconstruction methods for small artifacts", Remote Sens. Spatial Inf. Sci., XL-5/W4, pp.101-108, 2015. Doi: 10.5194/isprsarchives-XL-5-W4-101-2015.
- [12] E.Y. Putra, A.K. Wahyudi, and C. Dumingan, C, "A proposed combination of photogrammetry, augmented reality and virtual reality headset for heritage visualization", 2016 International Conference on Informatics and Computing, pp. 43-48, 2016.
- [13] N. Lercari, J. Schulze, W. Wendrich, B. Porter, M. Burton, and T.E. Levy, "3-D digital conservation of atrisk global cultural heritage", C.E. Catalano, L. De Luca (Eds), Eurographics Workshop on Graphics and Cultural Heritage, pp. 193-197, 2016. Doi:10.2312/gch.20161395.
- [14] B.J. Fernandez-Palacios, D. Morabito, and F. Remondino, "Access to complex reality-based 3D models using virtual reality solutions", Journal of Cultural Heritage, pp. 23, 40-48, 2017. Doi: 10.1016/j.culher.2016.09.003.
- [15] T.P. Kersten, G. Büyüksalih, F. Tschirschwitz, T. Kan, S. Deggim, T. Kaya, and A.P. Baskaraca, "The selimiye mosque of edirne, turkey – an immersive and interactive virtual reality experience using HTC vive". The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-5/W1, pp.361-367, 2017. Doi: 10.5194/isprs-archives-XLII-5-W1-403-2017.
- [16] M. Canciani, E. Conigliaro, M. Del Grasso, P. Papalini, and M. Saccone, "3D survey and augmented reality for cultural heritage. the case study of aurelian wall at Castra Praetoria in Rome", The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLI-B5, pp. 931-937, 2016. Doi: 10.5194/isprsarchives-XLI-B5-931-2016.
- [17] N. Snavely, S.M. Seitz, and R. Szeliski, "Modeling the world from internet photo collections". Int J Comput Vis, vol. 80, pp. 189-210, 2008. Doi: 10.1007/s11263-007-0107-3.
- [18] N. Snavely, and R. Szeliski, "Photo tourism: Exploring photo collections in 3D", ACM Transactions on Graphics, vol. 25, no. 3, pp. 835-846, 2006. Doi: 10.1145/1141911.1141964.
- [19] H. Lee, T.H. Jung, M.C.T. Dieck, and N. Chung, N, "Experiencing immersive virtual reality in museums", Information & Management, vol. 57, no. 5, 103229, 2020. Doi: 10.1016/j.im.2019.103229.
- [20] M. Trunfio, M.D. Lucia, S. Campana, and A. Magnelli, "Innovating the cultural heritage museum service model through virtual reality and augmented reality: the effects on the overall visitor experience and satisfaction", Journal of Heritage Tourism, vol. 17, no. 1, pp. 1-19, 2022. Doi: 10.1080/1743873X.2020.1850742.
- [21] W. Rosalee, W, "Teaching texture mapping visually", Retrieved May 26, 2023, from [https://users.cs.northwestern.edu/~ago820/cs351/Slides/r\\_wolfe\\_mapping.pdf](https://users.cs.northwestern.edu/~ago820/cs351/Slides/r_wolfe_mapping.pdf)



# Identifying and Prioritizing Digital Transformation Elements Using Fuzzy Analytic Hierarchy Process

Mohammed Hitham M.H, Hatem Elkadi, Neamat El Tazi

Faculty of Computers & Artificial Intelligence, Cairo University, Cairo, Egypt

**Abstract**—Digital transformation addresses multiple aspects of the organization. These aspects are the elements to be addressed for the digital transformation in any organization and are categorized as dimensions and sub-dimensions. In this work, these elements are collected from a wide range of related literature (56 publications). The most relevant elements were then identified through expert survey; involving 12 experts. The weights for these elements were identified using multi-criteria decision-making (MCDM) techniques. The Analytical Hierarchy Process (AHP) is one of the most often used MCDM techniques to incorporate individual and subjective preferences when conducting analysis and convert complex issues into a clear hierarchical structure. This work applies fuzzy AHP to take into consideration the treatment of uncertainty issues (in AHP), using the geometric mean method, and through an iterative process, calculate the weights of various dimensions and sub-dimensions, and prioritize them within the proposed roadmap for digital transformation implementation. Sensitivity analysis and comparison with AHP were used to validate our findings and the robustness of our approach. The proposed approach identified 9 main dimensions and 42 sub-dimensions which align with the majority of the literature. However, the advantage of this approach is the prioritization of these nine dimensions and their sub-dimensions as per the weights assigned to each one of them, allowing the project manager to allocate the available resources to the dimensions with the highest priority. The results show that the strategy and business process dimensions are the most crucial ones in the implementation of digital transformation.

**Keywords**—Digital transformation; MCDM; AHP; fuzzy AHP introduction

## I. INTRODUCTION

Digital Transformation (DT) has become an essential part of human life, and it is necessary for almost every private and public sector seeking growth, expansion, quality, and sustainability [1]. It can also change our life, work, increase productivity, save money, and reduce effort. In order to benefit from these advantages, several countries have started launching digital transformation projects such as [2-3]. Moreover, private sectors have also embraced DT, with organizations implementing their own DT programs [4]. However, it is important to note that every organization operates in a unique context and may be at a different stage of implementing DT. Therefore, it is essential for both public and private sectors to have an approach that allows them to assess their current position in DT implementation, identify strengths and weaknesses, and develop strategies to overcome any challenges [5]. By understanding their current state and addressing weaknesses, organizations can enhance their DT efforts and achieve greater success in embracing the advantages of DT.

This approach is called digital maturity model, readiness tests, or frameworks. It has two objectives: 1) the first objective is used to define the current position in the context of DT, and 2) the second objective is to propose a roadmap for implementation of DT. Organizations need a roadmap to clearly understand the DT concepts involved and effectively implement DT. The formulation of the roadmap poses a major challenge given the large variety of frequently occurring dimensions and sub-dimensions (criteria) that necessitate the use of a decision support technique called Multiple Criteria Decision Making (MCDM) [6]. To handle the large variation between the decision makers' opinions, Saaty [7] proposed AHP in order to streamline complex multi-decision-making processes and make them more systematic. AHP resolves complicated scenarios, including multiple criteria in the decision-making process by converting to a hierarchical structure [7-8]. Following the creation of the hierarchical structure, any two criteria are compared using pairwise comparison. There are three primary steps that make up the AHP: 1) define the goal and hierarchical structure of the study, 2) construct pairwise comparisons between criteria at each level of structure, and 3) calculate weight and ranking. AHP is the most widely used among MCDM techniques in domains, such as software [7] and industry [8]. Downsides with uncertainty associated with the decision-makers judgment can be solved by combining AHP and fuzzy set theory [10], [11-14].

To the best of our knowledge, the majority of studies look at how to evaluate digital transformation by defining the dimensions and sub-dimensions of digital transformation in the private or public sector and setting priorities for their implementation, but no study has taken more attention to a comprehensive approach that takes into account both. The study aims to address the gap in research by taking a comprehensive approach to evaluating DT in both the private and public sectors. It encompasses two key aspects: Firstly, the comprehensive synthesis of diverse elements, including dimensions and sub-dimensions, to DT within both the private and public domains. Secondly, the introduction of a hybrid approach—the combining of the Fuzzy Analytic Hierarchy Process (FAHP) with the Analytic Hierarchy Process (AHP)—designed to effectively prioritize the implementation of digital transformation components. The output of this prioritization will serve as the basis for a future roadmap proposal. Conducting such a study could help identify commonalities and differences between sectors, enabling a more effective allocation of resources and prioritization of implementation strategies. The rest of the paper is organized as follows: Section II presents a literature review of the relevant literature

on the topic. Section III discusses the research methodology employed in the study is discussed in detail; Section IV discusses the results of our approach. Section V validates the results of the hybrid approach by using sensitivity analysis and comparison with AHP, and in the finally section, the conclusion and future work are presented.

## II. LITERATURE REVIEWS

Selecting the appropriate maturity components, such as dimensions and sub-dimensions, and computing the weights requires an analytical and scientific approach, as follows in our work:

### A. Approaches to Weight Dimensions and Sub-Dimensions in DT

According to our literature review, there are two most common methods for defining weights for dimensions and sub-dimensions:

- The first method involves calculating the arithmetic mean.

In this procedure, specialists assign a separate value to each dimension and sub-dimension [15-18]. These values are then used to calculate mean values, which are considered as weights for each dimension and sub-dimension.

- The second method relies on MCDM

The second approach employs Multi-Criteria Decision Making (MCDM) techniques. In this method, experts assign comparative values to each dimension relative to the other dimensions. Likewise, they assign values to each sub-dimension relative to other sub-dimensions within the same dimension. Several studies have proposed various methodologies for prioritizing DT in different domains. For instance, [8] introduced an AHP-based approach for Industry 4.0, [13] employed Fermatean AHP for Supply Chain prioritization, [19] presented a method for technology selection in DT, [20] combined SF-AHP and SF-TODIM approaches in the defense industry, [21] devised a DEMATEL-based method for assessing DT in the health sector, [22] utilized ANP for evaluating DT in manufacturing, [23] introduced a fuzzy TOPSIS-based approach for supplier evaluation in DT within production systems and [24] employed Shannon entropy to calculate Business digital maturity in Europe. Analysis of previous research reveals that many studies focused on the private sector, and there is not the same level of interest in the public sector.

### B. Determining the DT Dimensions and Sub-dimensions

The literature review encompassed a thorough examination of assessment frameworks related to DT. This involved extracting DT maturity dimensions and their corresponding sub-dimensions from various studies [25-73]. The selection of these studies was based on their relevance to DT assessment requirements. The outcome of the literature review revealed a total of nine main dimensions and 168 corresponding sub-dimensions related to DT maturity. However, in order to streamline the assessment framework, only the most frequently occurring sub-dimensions, with a frequency of two or more,

were chosen. As a result, the sub-dimensions were reduced to a more manageable number of 70.

The results of the literature review to define dimensions and sub-dimensions can be summarized in Table I.

## III. RESEARCH METHODOLOGY

Based on a thorough study of the literature [25-73], including comparisons of digital maturity assessment in the field of DT and expert reviews. This research employs an iterative and tested approach to construct an assessment framework in DT [14-15] and [8]. Overall, research methodology has a two-phase process, namely:

- Defining dimensions and sub-dimensions in DT
- Derivation of weights via a hybrid approach (FAHP with AHP)

The output of phase one is used as an input to phase two. Each phase will be discussed as follows:

### A. Defining Dimensions and Sub-dimensions in DT

In Section II (B), drawing on the literature review, a first draft of the dimensions and sub-dimensions is defined. As described earlier, we need a way to identify the most relevant sub-dimensions for evaluating digital transformation. To achieve this, a review of the first draft with DT specialists (12) was conducted to capture the final relevant dimensions that were identified for further weight derivation. The summary of the methods used in this phase is shown in Fig. 1.

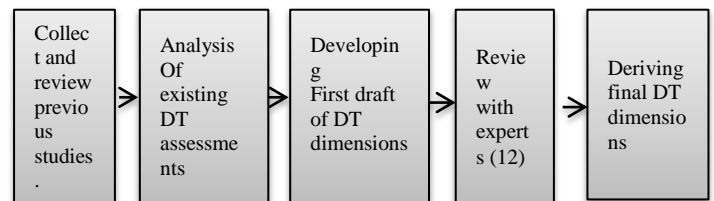


Fig. 1. Flow diagram of the first phase.

### B. Deriving Weights Using Hybrid Approach

As mentioned before, the aim of this research is to use the MCDM approach to prioritize the implementation of DT dimensions and sub-dimensions. This is done by proposing a hybrid approach that combines fuzzy group theory with the AHP method. Fig. 2 shows the proposed methodology. An overview of our approach will be given as follows:

- Step 1: Defining Problem and Planning: Define the objective of the study, define DT elements, and decompose the problem into a hierarchical structure.
- Step 2: Construct pair-wise comparisons at each level of the hierarchy structure by using fuzzy numbers.

The fuzzy scale used in the research [13] was employed to facilitate pairwise comparisons between DT elements, such as dimensions or sub-dimensions).

A is a  $n \times n$  pairwise matrix in which the relative importance of pairwise comparisons is determined on a scale of 1 to 9.

TABLE I. DT DIMENSIONS AND SUB-DIMENSIONS OF DT FROM LITERATURE REVIEWS

Dimensions Name	Sub-dimensions Name
Customer [25-29] [51]	customer experience [25-26], customer insight and analytics [27-28], competence with modern ICT [47-48], customer training [48-49], customer centricity [26][47][50] and customer integration[49-51]
Technology[25] [31-33] [36-38] [41-58] [59-61]	exploitation new technology AI, cloud computing, big data[25][36-38][56-61], IT architecture[31-33] [41-58], integration systems layer[25], Use technology for data collection [36-38] [41-58] [59-61], technology driven[31-33][42][59], digital capabilities[25][31-33], IT Infrastructure[31-33] [41-58], IT standard[33][36-38], effective technology planning[31-33][42][59], IT governance[25][31-33], define digital transformation requirements[61], and IT security[41-58].
Strategy [25-36][39] [45-46] [49] , [65-66]	coordination of digital transformation activities[25-30], strategic governance[26-30], technology investments[47][49][50], risk assessment for digital transformation[39-44], ecosystem management[60-64], stakeholder management[64-66] , strategic alignment [27][60][66],digital transformation vision[25-28][461-63], transformation in digital leadership[25-36], define role, Standards[62-64], top management commitment to realize digital transformation [47][49][50], and cost benefit analysis[45-46].
Organization [26-28] [31-32] [35-38] [45] [49] [52][57] [67-68],	organizational structure [26-28][31-32], organization collaboration[52][57], transformation in digital leadership[35][38][45], organization change management[31][32][57][68], cross functional collaboration [35-38][67], training[68], sufficient financial resources[32], and digital portfolio management[45].
Processes process[30][35-36] [39][42] [65] [69-70],	business process integration [30][35-36][39], business process performance management[42][69-70], business process standard[39][42] [65], business process security[30][42] [65], transformation in digital leadership[42][69-70], quality of business processes[39][69-70], Process control, intelligent process management[70]; reduce the costs of business process [42] and real-time insights & analytics [69-70].
Culture[26-27][31][42][45-46] [53-54]	Innovative culture [26-27], openness to change [26][42][45], communication[45-46][53], everyone is allowed to make decisions45-46][53-54], open environment[31] and digital education[53].
Data[26-27][33] [35] [48][50][65] [67]	data analysis [26-27][33] [35] [48][50][65] [67], data management[33] [35] [48][50], data security and privacy[26-27][33] [35], data governance[[33][48][67], data quality[67], data visualization[33][65] and data archiving[48].
Employee[26-27][42][45] [52][72]	Openness to new technology [26-27] [42] [45], willingness to change [52] [72] and employee training. [26-27] [42][45] [52] [72].
Citizen [72-73].	Citizen training [72-73], citizen skills [72-73] and citizen centricity [72-73].

$$\tilde{A}^k = \begin{bmatrix} \tilde{d}_{11}^k & \tilde{d}_{12}^k & \dots & \tilde{d}_{1n}^k \\ \tilde{d}_{21}^k & \dots & \dots & \tilde{d}_{2n}^k \\ \dots & \dots & \dots & \dots \\ \tilde{d}_{n1}^k & \tilde{d}_{n2}^k & \dots & \tilde{d}_{nn}^k \end{bmatrix} \quad (1)$$

Where  $\tilde{d}_{ij}^k$  indicates the Jth decision maker's preference of ith criterion over the jth criterion, via fuzzy a triangular numbers. It is fuzzy number (l, m, u) [13], for reciprocal:

$$d_{ij} \tilde{k}^{-1} = (l, u, m)^{-1} = \left( \frac{1}{u}, \frac{1}{m}, \frac{1}{l} \right) \quad (2)$$

Twelve decision makers "experts" consist of DT consultants and academics, which are considered experts in their respective fields abbreviated as E1, E2, E12. By collecting the opinions of these decision-makers and constructing the pairwise comparison matrix using Eq. (1), it becomes possible to determine the relative of each dimension and sub-dimension Pairwise comparisons of the fuzzy judgment matrix "i" are frequently inconsistent because they are prone to bias and inaccuracy in preference for expert responses. Therefore, AHP is used to avoid inconsistencies in responses. The consistency index for pairwise comparisons was calculated by using Eq. (3).

$$\lambda_{max} = \frac{1}{n} \sum_{j=1}^n \frac{A_{wi}}{w_i} \quad (3)$$

$$CI = \frac{\lambda_{max} - n}{n - 1} \quad (4)$$

Where n is the number of dimensions or sub- dimensions

Eq. (5) is used to calculate the consistency ratio, where CI is compared with a random index.

$$CR = \frac{CI}{RI} \quad (5)$$

- Step 3: Check consistencies (for the most likely value)

This random index (RI) value [12] is correlated to the number of dimensions or sub-dimensions compared and used to calculate the consistency ratio, as shown in Eq. (5). The level of consistency is acceptable if the CR is less than 0.1. If not, there will likely be a lot of inconsistency, so the opinion of the decision-maker will be deleted. In this study, the CI is calculated for the middle value (most likely value "m") [11], even though the pairwise comparison indices (relative importance) of the judgment matrix are TFNs for each decision-maker separately. In this work, we calculate the consistency ratio for each expert separately. If the consistency index exceeds 0.1, the opinion of this expert will be deleted.

- Step 4: Aggregate expert opinions

If there are many decision makers accepted  $\tilde{d}_{ij}$ , the average"  $\tilde{d}_{ij}$  "is calculated using Eq. (6) [13].

$$\tilde{d}_{ij} = \frac{\sum_{k=1}^k \tilde{d}_{ij}^k}{k} \quad (6)$$

According to averaged preferences, pair wise contribution matrices are updated as shown in Eq. (7).

$$\tilde{A} = \begin{bmatrix} \tilde{d}_{11}^k & \dots & \tilde{d}_{1n}^k \\ \vdots & \ddots & \vdots \\ \tilde{d}_{n1}^k & \dots & \tilde{d}_{nn}^k \end{bmatrix} \quad (7)$$

- Step 5: Calculate CR to Aggregate Expert Opinions

Pair-wise comparisons were constructed for the opinions of decision-makers based on Eq. (6), and then a new CR was calculated for this matrix using Eq. (5).

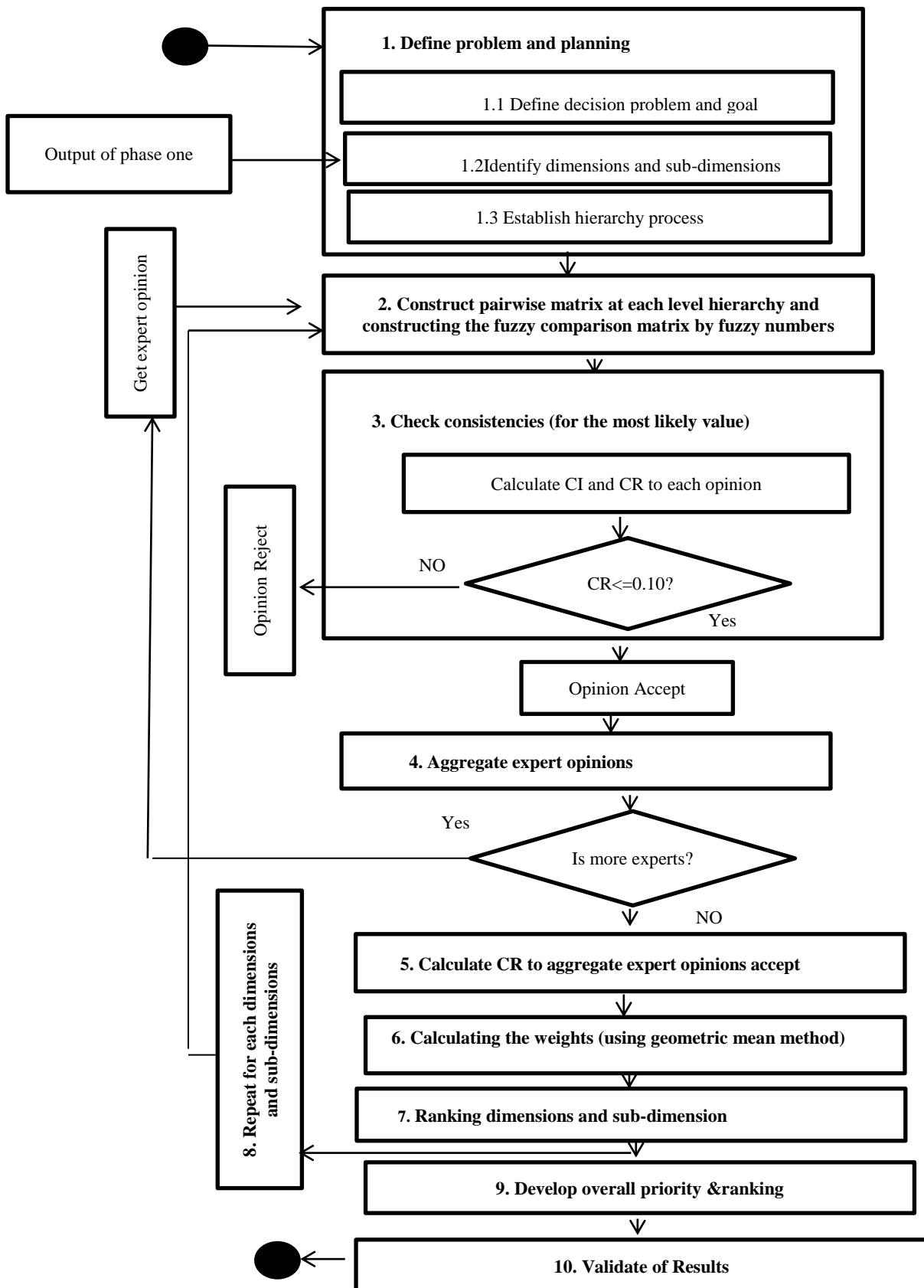


Fig. 2. Flowchart of the proposed methodology of the second phase.

- Step 6: Calculating the weights Using Fuzzy Geometric Mean Method.

As mentioned before, the major problem of AHP has been enhanced by utilizing fuzzy logic since it does not include vagueness for subjective judgments. There are several approaches to F-AHP, such as the geometric mean [11], [14], [64-65], and the extent analysis method [66]. In this work, the fuzzy geometric mean method was used to calculate the weights.

The sixth step contains several sub-steps that can be summarized as follows:

Step 6.1: According to [64] and [13], the fuzzy geometric mean value of each sub-dimension or dimension is calculated using Eq. (8). Here  $\tilde{r}_i$ , it still represents triangular values.

$$\tilde{r}_i = \left( \prod_{j=1}^n \tilde{d}_{ij} \right)^{1/n}, i = 1, 2, \dots, n \quad (8)$$

Where n is the number of dimensions or sub-dimensions.

Step 6.2: Find the vector summation of each  $\tilde{r}_i$ .

Step 6.3: Find the (-1) power of the summation vector. Replace the fuzzy triangular number, to make it in an increasing order [13].

Step 6.4: The fuzzy weight of dimensions or sub-dimensions was calculated as shown in Eq. (9).

$$\tilde{w}_i = \tilde{r}_i \otimes (\tilde{r}_1 \oplus \tilde{r}_2 \oplus \dots \oplus \tilde{r}_n)^{-1} = (lw_i, Lw_i, mw_i, uw_i) \quad (9)$$

Step 6.4: The weights that have been calculated by using Eq. (8) are still fuzzy triangular numbers, so we need to defuzzified them by the Centre of Area (COA) as shown in Eq. (10) [13].

$$W_i = \frac{l+m+u}{3} \quad (10)$$

Step 6.5: The weights that come from Eq. (9) were normalized as shown in Eq. (11).

$$N_i = \frac{M_i}{\sum_{i=1}^n M_i} \quad (11)$$

- Step 7: Ranking of dimensions and sub-dimensions.

Based on the outputs of step seven, the dimensions and sub-dimensions can be ranked according to weights.

- Step 8: Repeat steps 3, 4, and 5 for all levels of the hierarchy.
- Step 9: Develop overall priority & ranking.

According to [9], the total weight of sub-dimensions can be calculated according to Eq. (12), where “I” is the weight of dimensions and “j” is the weight of sub-dimensions in each dimension.

$$t_{ij} = g_i \times w_{ij} \quad (12)$$

- Step 10: Validate of Results

The Sensitivity analysis and comparison with AHP were used to validate of our approach.

#### IV. RESULTS AND DISCUSSION

In this section, we used the proposed method presented in Section III, as illustrated in in Fig. 1 and 2 to define and prioritize the dimensions and sub-dimensions of digital transformation. It will be discussed as follows:

##### A. DT Dimensions and Sub- dimensions

The results of the review with experts (applying Method 1 in Fig. 1) to define relevant dimensions and sub-dimensions are summarized in Table II. After conducting the review with experts to determine the most important sub-dimensions in evaluating digital transformation, the sub-dimensions were reduced to 42.

TABLE II. DT DIMENSIONS AND SUB-DIMENSIONS AFTER REVIEW WITH EXPERTS

Dimensions	Sub-dimensions
Customer	Customer training, Customer centricity, Customer integration
Technology	IT Architecture, Technology driven, Technologysecurity ,IT governance, Exploitation new technology,Use technology for data collection ,Digital Capabilities, IT Infrastructure, IT standard, Effective technology planning
Strategy	Coordination of digital transformation activities, Strategic governance, Technology investments, Risk assessment for digital transformation , Ecosystem Management, Stakeholder Management, Strategic alignment (Business-IT alignment) ,Digital transformation vision, Transformation in Digital Leadership
Organization	Transformation in digital leadership, Organization governance, Digital change management, Cross functional collaboration
Processes process	Business process Integration, Business process performance management , Business process standard, Business process security , Transformation in digital leadership
Culture]	Innovative culture, Openness to change,Communication, Everyone is allowed to make decisions
Data	Data analysis, Data management ,Data security and privacy,Data governance
Employees	Openness to new technology, Willingness to change, Employee training
Citizen	Citizen training, Citizen skills, Citizen centricity

##### B. Weights of DT Dimensions and sub-dimensions

In this section, the proposed method presented in Section III in Fig. 2 is used to prioritize the implementation of the dimensions and sub-dimensions of digital transformation by calculating weights. Fig. 3 illustrates this hierarchical structure involving the objective of the study, dimensions, and sub-dimensions.

In Fig. 3, the first level relates to the objective goal of the study. The second level corresponds to dimensions, and the last level corresponds to sub-dimensions of each dimension. In this paper, a pairwise comparison matrix will be created between elements (dimensions) in level 2. Similarly, a pairwise comparison matrix will be created between elements (sub-dimensions) in level 3 that have the same parent in level 2. Due to space constraints, the results of the steps involved in the proposed method are presented for the main dimensions on level 2, as shown in subsection A.

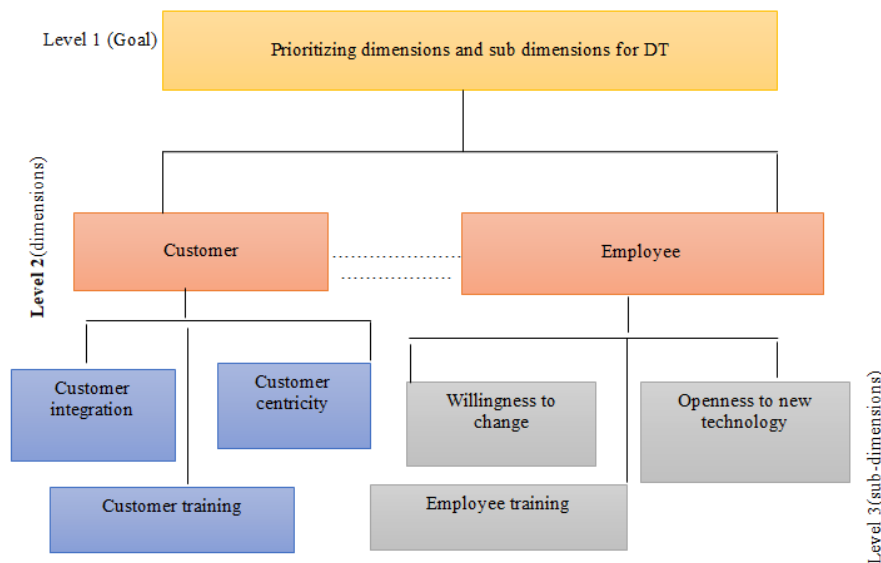


Fig. 3. Example DT Hierarchy structure of the problem.

### C. Weights of DT Dimensions

The results of the steps involved in the proposed method are presented for the pairwise comparison matrix between the main dimensions on level 2, as well as the output of each step. For each expert out of 12, a pairwise matrix was created, but due to the difficulty of displaying all of them, it was sufficient to present a matrix for one expert, as shown in Table III. Then consistency is checked for each expert (E) separately, as shown in Table IV. Only four expert opinions were accepted, while eight expert opinions were omitted, as shown in Table IV. The opinions of the experts accepted in the previous step were collected, as shown in Table V. The consistency ratio of the opinions of the accepted experts is calculated based on Table V. Consistency ratio = 0.064003. The outputs of applying Eq. (8), (9), and (10) and Step 7 are summarized in Table VII. From Table VI, it can be noticed that the strategy dimension has the highest weight (priority) "0.341" followed by the business process with a weight "0.215". Thus, the strategy dimension will rank first, followed by the business process. It can also be seen that the citizen dimension has the least weight (0.030).

### D. Weights of Sub-dimensions

As previously mentioned, due to space limitations, the results of the steps involved in the proposed method will not be presented for the main dimensions at Level 3, but the final results for the respective weights for each sub-dimension will be shown in Table VII.

### E. Total Weights of each Sub-dimensions

As we mentioned before, the total overall weight of each sub-dimension (t) can be calculated according to Eq. (12). The results of applying step eight can be summarized in Table VIII. For example, in "digital transformation vision",  $g_i=0.341$ ,  $w_{ij}=0.355$ , so  $t_{ij}=0.121$ . After calculating  $t_{ij}$  for all, it can be ranked. Based on the outputs of Table VII, it is possible to arrange the implementation of the sub-dimensions in relation to digital transformation. "Digital transformation vision" was first ranked, "business process standard" was placed second, and

"integration of citizens" came in last ranked. So it can be said that "digital transformation vision" is the leading factor for DT, followed by "business process standard". One other salient sub-dimension is willingness of employees to change" followed by "business-IT alignment". The consistency analysis of this research is summarized in Table VIII. Fig. 4 shows an incremental comparison of the total weights of all sub-dimensions (t) in detail.

## V. RESULTS VALIDATIONS

In this section, our work will be evaluated by identifying the advantages of this work compared to the research that is most similar to it [8] and comparing the results of our work with the results of AHP, in addition to using the Sensitive Analysis.

### A. Comparison with Prior Study

In order to contextualize our research, it's imperative to draw comparisons with a prior study [8]. This prior research shares the commendable attribute of employing a coherent methodology to delineate and assign weights to DT elements. Nonetheless, the preceding study harbors three notable limitations: it confines its focus solely on the private sector for the definition of DT elements, employs the AHP to prioritize these elements despite inherent uncertainties, and regrettably omits result validation. In response to these challenges, this research endeavors to address them comprehensively. The first limitation was overcome by the comprehensive identifying of elements relevant to DT evaluation in general (both segments). The second challenge is strategically navigated by adopting a combined approach, unifying AHP with FAHP to bolster consistency and mitigate the uncertainties often associated with expert judgments. Furthermore, a rigorous sensitivity analysis was performed to validate the results, critically addressing the last limitation. In doing so, our research not only endeavors to provide a comprehensive solution but also contributes to the broader scholarly discourse on digital transformation assessment methodologies.

**B. Comparison Results (Ranking)**

A comparative analysis is performed to validate the effectiveness of our proposed approach by comparing the results of our approach with those of AHP as follows:

- Comparing the ranking between the main dimensions

Based on the results obtained, it can be observed that the ranking of the main dimensions using the Analytic Hierarchy Process (AHP) is the same as the ranking using fuzzy AHP, with the exception of the business process and employee dimensions as shown in Fig. 5. In AHP, the business process dimension is ranked third, whereas in fuzzy AHP, it is ranked second. Similarly, the employee dimension is ranked second in AHP and third in fuzzy AHP. Comparative analysis of the

results indicates that our approach is 80% compatible with AHP in terms of dimensional order. This suggests that there is a significant level of agreement between the two methods, except for the specific dimensions mentioned above.

- Comparing the ranking between the sub-dimensions in each dimension

Due to space limitations, only the sub-dimensions rank of the data dimension was compared. Based on the comparative results shown in Fig. 6, it can be concluded that our approach is 100% compatible with AHP in terms of the ordering of sub-dimensions in the data dimension. This indicates that our proposed approach accurately orders the implementation of dimensions in the decision tree (DT).

TABLE III. FUZZIFIED PAIRWISE MATRIX BETWEEN DIMENSIONS FOR FIRST EXPERT

	<i>Strategy</i>	<i>Business process</i>	<i>Employee</i>	<i>Data</i>	<i>Technology</i>	<i>Organization</i>	<i>Stakeholder</i>	<i>Culture</i>
Strategy	(1,1,1)	(1,1,1)	(4,5,6)	(6,7,8)	(1,1,1)	(4,5,6)	(2,3,4)	(2,3,4)
Business process	(1,1,1)	(1,1,1)	(2,3,4)	(6,7,8)	(4,5,6)	(6,7,8)	(6,7,8)	(6,7,8)
Employee	(01.6,2,25)	(0.25,0.33,0.5)	(1,1,1)	(2,3,4)	(6,7,8)	(6,7,8)	(6,7,8)	(6,7,8)
Data	(0.12,0.14,0.16)	(0.12,0.14,0.16)	(0.25,0.33,0.5)	(1,1,1)	(2,3,4)	(2,3,4)	(2,3,4)	(2,3,4)
Technology	(1,1,1)	(0.16,2,0.25)	(0.12,0.14,0.16)	(0.25,0.33,0.5)	(1,1,1)	(1,1,1)	(4,5,6)	(4,5,6)
Organization	(01.6,0.2,0.25)	(0.12,0.14,0.16)	(0.12,0.14,0.16)	(0.25,0.33,0.5)	(1,1,1)	(1,1,1)	(4,5,6)	(4,5,6)
Stakeholder(customer or citizen)	(0.25,0.33,0.5)	(0.12,0.14,0.16)	(0.12,0.14,0.16)	(0.25,0.33,0.5)	(0.25,0.33,0.5)	(01.6,0.2,0.25)	(1,1,1)	(1,1,1)
Culture	(01.6,2,25)	(0.12,0.14,0.16)	(0.12,0.14,0.16)	(1,1,1)	(1,1,1)	(01.6,0.2,0.25)	(0.25,0.33,0.5)	(0.25,0.33,0.5)

TABLE IV. CHECK CONSISTENCY FOR EACH EXPERT

<i>Expert #</i>	<i>CR</i>	<i>Decision (Accept or Reject)</i>
E1	0.07	Accept
E2	0.20	Reject
E3	0.19	Reject
E 4	0.03	Accept
E 5	0.06	Accept
E 6	0.25	Reject
E 7	0.16	Reject
E 8	0.02	Accept
E 9	0.13	Reject
E 10	0.11	Reject
E 11	0.10	Reject
E 12	0.16	Reject

TABLE V. FUZZIFIED PAIRWISE MATRIX BETWEEN DIMENSIONS FOR ACCEPT OPINIONS

	<i>Strategy</i>	<i>Business process</i>	<i>Employee</i>	<i>Data</i>	<i>Technology</i>	<i>Organization</i>	<i>stakeholder</i>	<i>Culture</i>
Strategy	(1,1,1)	(1,1,1)	(4,5,6)	(2.44,2.64,2.82)	(1,1,1)	(4.89,5.91,6.92)	(4.24,5.19,6)	(4.24,5.19,6)
Business process	(1,1,1)	(1,1,1)	(4.24,5.19,6)	(4.89,5.91,6.92)	(4,5,6)	(6,7,8)	(6,7,8)	(6,7,8)
Employee	(0.16,2,0.25)	(0.16,0.19,0.23)	(1,1,1)	(1.41,1.73,2)	(3.46,4.58,5.66)	(3.46,4.58,5.66)	(2.44,2.64,2.82)	(2.44,2.64,2.82)
Data	(0.35,0.37,40)	(0.14,0.16,0.20)	(0.5,0.57,0.70)	(1,1,1)	(1.41,1.73,2)	(3.46,4.58,5.65)	(3.46,4.58,5.66)	(3.46,4.58,5.66)
Technology	(1,1,1)	(0.16,2,0.25)	(0.17,0.21,0.28)	(0.5,0.57,0.70)	(1,1,1)	(2.44,2.64,2.82)	(4.89,5.91,6.98)	(4.89,5.91,6.98)
Organization	(0.14,0.16,0.20)	(0.12,0.14,0.16)	(0.17,0.21,0.28)	(0.17,0.21,0.28)	(0.35,0.37,40)	(1,1,1)	(2,2,2,2.44)	(2,2,2,2.44)
Stakeholder**customer or citizen	(0.16,0.19,0.23)	(0.12,0.14,0.16)	(0.35,0.37,40)	(0.17,0.21,0.28)	(0.14,0.16,0.20)	(0.40,0.45,0.5)	(1,1,1)	(1,1,1)
Culture	(0.16,0.19,0.23)	(0.12,0.14,0.16)	(0.35,0.37,40)	(0.17,0.21,0.28)	(0.14,0.16,0.20)	(0.40,0.45,0.5)	(1,1,1)	(1,1,1)

TABLE VI. WEIGHT OF DIMENSIONS USING GEOMETRIC MEAN

<i>Dimension Name</i>	<i>Fuzzy wi</i>	<i>Centre of Area (COA)</i>	<i>Normalized wi</i>	<i>Rank</i>
Strategy	0.251,0.352,0.455	0.353	0.341	1
Business process	0.167,0.223,0.277	0.222	0.215	2
Employee	0.093,0.136,0.182	0.137	0.132	3
Data	0.070,0.097,0.216	0.128	0.123	4
Technology	0.064,0.084,0.108	0.085	0.083	5
Organization	0.032,0.043,0.057	0.044	0.042	6
Customer	0.025,0.033,0.045	0.034	0.033	7
Culture	0.024,0.033,0.045	0.033	0.030	9
Citizen	0.022,0.031,0.043	0.032	0.031	8

TABLE VII. WEIGHTING AND RANKING OF DT DIMENSIONS AND SUB-DIMENSIONS

<i>Dimensions Name</i>	<i>Weights of dimensions (g)</i>	<i>Sub-dimensions Name</i>	<i>Weights of sub-dimensions (w)</i>	<i>Total Weights (g*w)</i>	<i>Ranking Sub-dimensions</i>
Strategy	0.341	Digital transformation vision	0.355	0.121	1
		Coordination of digital transformation activities	0.243	0.082	4
		Business-IT alignment	0.154	0.052	7
		Technology investments	0.084	0.028	10
		Governance	0.072	0.024	11
		Ecosystem Management	0.039	0.013	20
		Stakeholder Management	0.029	0.0098	24
		Risk assessment for digital transformation	0.023	0.0078	29
Business process	0.215	Business process standard	0.55	0.1183	2
		Business process performance management	0.26	0.0559	6
		Business process Integration	0.14	0.0301	9
		Business process security	0.06	0.0129	21



Employee	0.132	Willingness to change	0.7380	0.0974	3
		Openness to new technology	0.1680	0.0222	12
		Employee training	0.0940	0.0124	22
Data	0.123	Data Analysis	0.5050	0.0621	5
		Data management	0.2750	0.0338	8
		Data security	0.1380	0.0170	17
		Data governance	0.0820	0.0101	23
Technology	0.083	technology planning	0.2320	0.0193	14
		Exploitation new technology	0.1180	0.0098	25
		Technology driven	0.1120	0.0093	26
		Technology security	0.1070	0.0089	27
		IT Infrastructure	0.0920	0.0076	30
		IT Architecture	0.0870	0.0072	31
		Use technology for data collection	0.0690	0.0057	36
		Digital Capabilities	0.0710	0.0059	35
		IT standards	0.0620	0.0051	38
		IT governance	0.0500	0.0042	39
Organization	0.0420	Cross functional collaboration	0.4410	0.0185	15
		Change management	0.3200	0.0134	19
		Organizational governance	0.1500	0.0063	33
		Transformation in digital leadership	0.0890	0.0037	41
Customer	0.033	Customer centricity	0.5160	0.0170	16
		Customer training	0.1950	0.0064	32
		Customer integration	0.1560	0.0051	37
Culture	0.031	Innovative culture	0.5300	0.01643	18
		Openness to change	0.2700	0.00837	28
		communication	0.1300	0.00403	40
		make decisions	0.0800	0.00248	42
Citizen	0.030	Citizen training	0.7340	0.02202	13
		Citizen centricity	0.1980	0.00594	34
		Citizen integration	0.0660	0.00198	43

TABLE VIII. CONSISTENCY RATIO OF AHP MATRICES

<i>Dimensions</i>	<i>Consistency Ratio (CR)</i>
Strategy	0.069238
Business process	0.069564
Employee	0.090259
Data	0.06956
Technology	0.025044
Organization	0.021964
Customer	0.088015
Culture	0.069564
Citizen	0.089011
Overall Consistency of Dimensions	0.064003

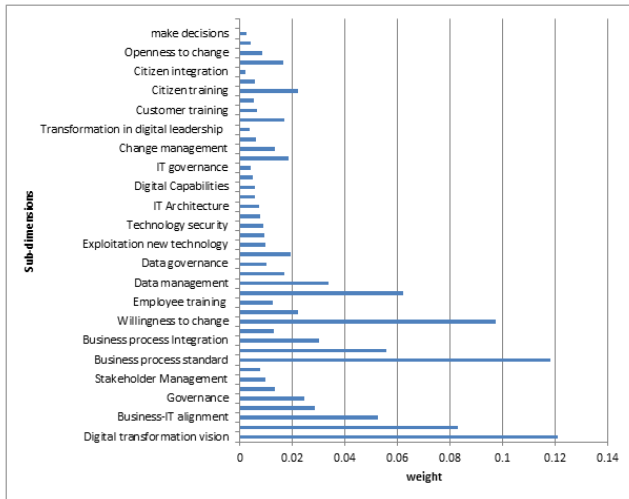


Fig. 4. Weight comparison of DT sub-dimensions

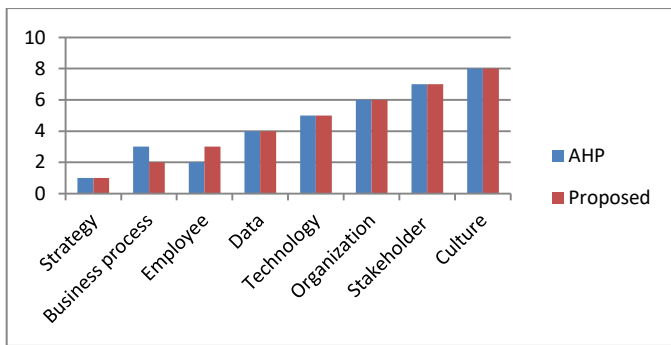


Fig. 5. Comparison results of the ranking of dimensions based on several evaluation approaches

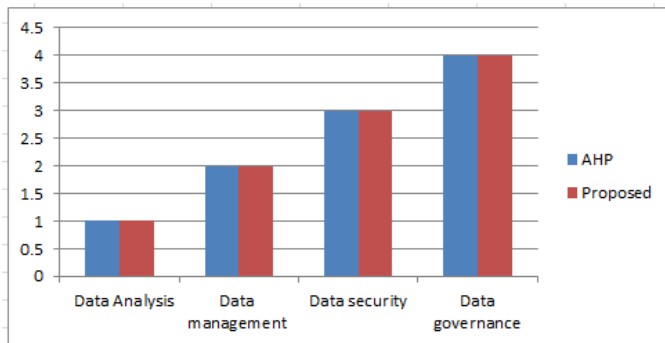


Fig. 6. Comparison results of the ranking data sub-dimensions based on several evaluation

### C. Sensitivity Analysis

A sensitivity analysis is a tool to determine the effects of potential modifications in the dimension or sub-dimension weights on the prioritization of DT [13]. A sensitivity analysis was applied to the FAHP approach results based on dimensions. The X-axis represents the change in important values between 1 and 9 (that have been assigned by 12 experts) of the main dimensions or sub-dimensions, and the Y-axis represents the ranking of dimensions. We can observe the effects on the ranking of the dimensions and sub-dimensions as follows:

- Sensitive analysis in dimensions

In this analysis, the weights of a certain dimension for each expert will be changed between 1 and 9, while the weights of other dimensions are fixed. For example, when the weight of the strategy dimension with respect to the business process dimension is changed between 1 and 9, strategy has always been placed in the first rank, except for one time when business process came first, as shown in Fig. 7. This will be iterated by changing the strategic dimension values for each of the remaining dimensions. By conducting a sensitivity analysis, it was determined that the weights assigned to the primary dimension have only a slight impact on the overall results. Additionally, the order of choices does not change significantly even with variations in the weights of the primary dimensions.

- Sensitive analysis in sub-dimensions (customer as example)

Due to space constraints, only sensitivity in customer sub-dimensions was examined, as shown in Fig. 8, 9, and 10.

- ✓ Sensitive analysis in customer training with respect to the customer centricity

When the weight of the customer training with respect to the customer centricity is changed, the customer training has always been placed in the first rank and the customer centricity has always been placed in the second rank except one time, as shown in Fig. 8.

- ✓ Sensitive analysis in customer training with respect to the customer integration

When the weight of the customer training with respect to the customer integration is changed, the customer training has always been placed in the first rank and the customer integration in the second rank, as shown in Fig. 9.

- ✓ Sensitive analysis in customer centricity with respect to the customer integration

When the weight of customer centricity with respect to customer integration is changed, customer centricity has always been placed in the second rank, as shown in Fig. 10.

Sensitivity analysis shows that weights for the customer sub-dimensions have only a limited effect on the results, and there is no significant change in the order of the sub-dimensions.

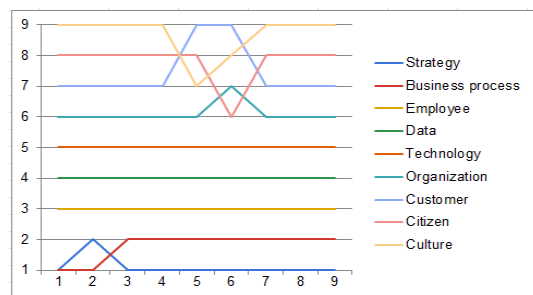


Fig. 7. Results of sensitivity analysis strategy dimension with respect to the technology dimension.

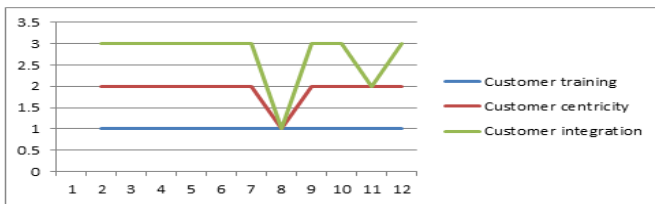


Fig. 8. Results of sensitivity analysis customer training sub-dimension with respect to the customer centricity sub-dimension.

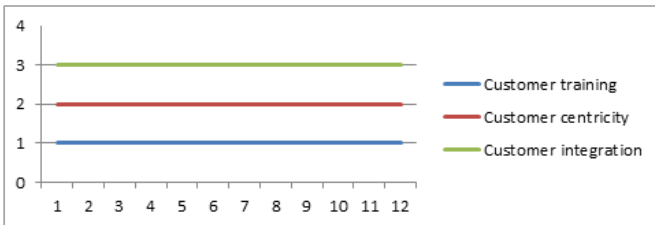


Fig. 9. Results of sensitivity analysis Customer training sub-dimension with respect to the Customer integration sub-dimension

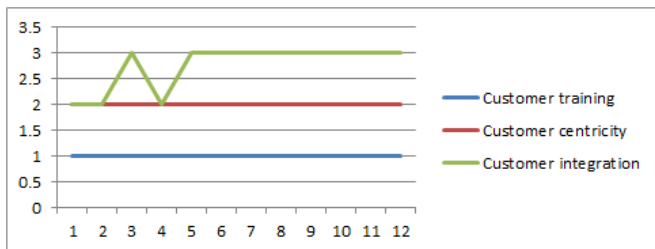


Fig. 10. Results of sensitivity analysis Customer centricity sub-dimension with respect to the Customer integration sub-dimension.

## VI. CONCLUSION

The core objective of this study was to establish a systematic framework for prioritizing the implementation of dimensions and sub-dimensions within the context of digital transformation. This was achieved through two distinctive phases. The initial phase involved defining the key dimensions and sub-dimensions, drawing from prior research and expert evaluations. A comprehensive set of 42 sub-dimensions was assembled under nine primary dimensions. Subsequently, the study progressed into the second phase, where the weights of both main dimensions and sub-dimensions were meticulously computed. In this research, the integration of the fuzzy geometric mean method with AHP provided the basis for identifying priority areas of focus for organizations. The application of the fuzzy scale and geometric mean method to allocate weights to dimensions and sub-dimensions effectively handled uncertainties in the decision-making process. The inclusion of AHP further bolstered decision consistency. The findings underscored that "strategy" (0.341) and "business process" (0.215) emerged as the two pivotal dimensions within the realm of digital transformation. The sub-dimension "digital transformation vision" held the foremost position, closely trailed by "business process standard. This study carries significant implications for organizational decision-makers across both the private and public sectors. It offers a tangible pathway for identifying the priority of sub-dimensions, thereby

amplifying the likelihood of successful digital transformation endeavors. Sensitivity analysis was then employed to validate the outcomes of our approach. Notably, the ranking of alternatives remained largely unchanged even when the weights of primary dimensions or sub-dimensions were modified. Furthermore, a comparative analysis was executed between our proposed approach and AHP. Through sensitivity analysis and consistency ratio calculations, the robustness and effectiveness of our approach were both established. In summation, this research introduces a methodological paradigm that guides the strategic sequencing of dimensions and sub-dimensions in digital transformation initiatives. It not only empowers decision-makers but also underscores the reliability and effectiveness of the proposed approach through rigorous analysis and validation.

- Limitation

The hybrid approach used in this paper was created exclusively for digital transformation. As well, this is a general approach and does not apply to case studies.

- Future work

Several experiments will be carried out using different MCDM techniques as well as applying our approach in many areas. The next step will be to use these findings to suggest a roadmap for the organizations when they are being evaluated.

## REFERENCES

- [1] Gebayew, Chernet, et al. "A systematic literature review on digital transformation." 2018 International Conference on Information Technology Systems and Innovation (ICITSI). IEEE, 2018
- [2] Gong, Yiwei, Jun Yang, and Xiaojie Shi. "Towards a comprehensive understanding of digital transformation in government: Analysis of flexibility and enterprise architecture." *Government Information Quarterly* 37.3 (2020): 101487.
- [3] de Lemos Santos, Francisco Miguel, et al. "ticAPP-Digital Transformation in the Portuguese Government." *ICEIS* (2). 2019.
- [4] Masuda, Yoshimasa, and Murlikrishna Viswanathan. *Enterprise architecture for global companies in a digital it era: adaptive integrated digital architecture framework (AIDAF)*. Springer, 2019.
- [5] Roza Carreño, Daniel Felipe. *An enterprise architecture framework for digital transformation*. MS thesis. University of Twente, 2020.
- [6] Vaidya OS, Kumar S (2006) Analytic hierarchy process: an overview of applications. *Eur J Oper Res* 169:1–29
- [7] Botchway, Ivy Belinda, et al. "Evaluating software quality attributes using analytic hierarchy process (AHP)." *International Journal of Advanced Computer Science and Applications* 12.3 (2021).
- [8] ETKESER, Sadi, and Lütfi APİLİOĞULLARI. "Designating Industry 4.0 Maturity Items and Weights for Small and Medium Enterprises." *Bilişim Teknolojileri Dergisi* 14.1 (2021): 79-86.
- [9] PFATB, Data. "Fuzzy Analytical Hierarchy Process (FAHP) using geometric mean method to select best processing framework adequate to big data." *Journal of Theoretical and Applied Information Technology* 99.1 (2021): 207-226.
- [10] Tesfamariam, Solomon, and Rehan Sadiq. "Risk-based environmental decision-making using fuzzy analytic hierarchy process (F-AHP)." *Stochastic Environmental Research and Risk Assessment* 21 (2006): 35-50.
- [11] Nabeeh, Nada A., et al. "Neutrosophic multi-criteria decision making approach for iot-based enterprises." *IEEE Access* 7 (2019): 59559-59574.
- [12] Ayhan, Mustafa Batuhan. "A fuzzy AHP approach for supplier selection problem: A case study in a Gear motor company." *arXiv preprint arXiv:1311.2886* (2013).

- [13] Alkan, Nursah, and Cengiz Kahraman. "Prioritization of supply chain digital transformation strategies using multi-expert Fermatean fuzzy analytic hierarchy process." *Informatica* (2022): 1-33.chain.
- [14] J. Becker, R. Knackstedt, J. Pöppelbuß, "Developing Maturity Models for IT Management", *Bus Inf Syst Eng*, 1, 213–222, 2009.
- [15] Huynh, Vy Dang Bich, et al. "Application of fuzzy analytical hierarchy process based on geometric mean method to prioritize social capital network indicators." *International Journal of Advanced Computer Science and Applications* 9.12 (2018): 182-186.
- [16] Chang, D.-Y., (1996) "Applications of the extent analysis method on fuzzy AHP", *European By applying the our approach , the dimensions and sub-dimensions can be prioritized and a descending-order to perform Project*.
- [17] Schumacher, Andreas, Selim Erol, and Wilfried Sihh. "A maturity model for assessing Industry 4.0 readiness and maturity of manufacturing enterprises." *Procedia Cirp* 52 (2016): 161-166.
- [18] benchmark digital maturity of the dutch health insurance companies with application of a PCA-model Niels P. Theunissen - 2016.
- [19] Erbay, Hasan, and Nihan Yıldırım. "Technology selection for digital transformation: a mixed decision making model of AHP and QFD." *Proceedings of the International Symposium for Production Research 2018 18*. Springer International Publishing, 2019.
- [20] Nebati, Emine Elif, Berk Ayyaz, and Ali Osman Kusakci. "Digital transformation in the defense industry: A maturity model combining SF-AHP and SF-TODIM approaches." *Applied Soft Computing* 132 (2023): 109896.
- [21] Erdal, Berrak, et al. "Digital Maturity Assessment Model Development for Health Sector." *Digitizing Production Systems*. Springer, Cham, 2022. 131-147.
- [22] Lee, Jeongcheol, et al. "A smartness assessment framework for smart factories using analytic network process." *Sustainability* 9.5 (2017): 794.
- [23] Aydın, Serhat, Ahmet Aktas, and Mehmet Kabak. "Evaluation of suppliers in the perspective of digital transformation: a spherical fuzzy TOPSIS approach." *Intelligent and Fuzzy Techniques for Emerging Conditions and Digital Transformation: Proceedings of the INFUS 2021 Conference, held August 24-26, 2021*. Volume 2. Springer International Publishing, 2022.
- [24] Tutak, Magdalena, and Jarosław Brodny. "Business digital maturity in Europe and its implication for open innovation." *Journal of Open Innovation: Technology, Market, and Complexity* 8.1 (2022): 27.
- [25]
- [26] ETKESER, Sadi, and Lütfi APİLİOĞULLARI. "Designating Industry 4.0 Maturity Items and Weights for Small and Medium Enterprises." *Bilişim Teknolojileri Dergisi* 14.1 (2021): 79-86.
- [27]
- [28] Buckley, J. J., (1985) "Fuzzy hierarchical analysis", *Fuzzy Sets Systems*, Vol.17 (1), 233–247.
- [29] Canetta, L., Barni, A., Montini, E.: Development of a digitalization maturity model for the manufacturing sector. In: 2018 Ieee International Conference on Engineering, Technology and Innovation (ICE/ITMC), pp. 1-7 (2018). IEEE
- [30] [29] Schumacher, A., Erol, S., Sihh, W.: A maturity model for assessing industry 4.0 readiness and maturity of manufacturing enterprises. *Procedia Cirp* 52, 161-166 (2016)
- [31] Schumacher, A., Nemeth, T., Sihh, W.: Roadmapping towards industrial digitalization based on an industry 4.0 maturity model for manufacturing enterprises. *Procedia Cirp* 79, 409-414 (2019)
- [32] Valdez-de-Leon, O.: A digital maturity model for telecommunications service providers. *Technology innovation management review* 6(8) (2016)
- [33] Goumeh, F., Barforoush, A.A.: A digital maturity model for digital banking revolution for iranian banks. In: 2021 26th International Computer Conference, Computer Society of Iran (CSICC), pp. 1-6 (2021). IEEE
- [34] Pirola, F., Cimini, C., Pinto, R.: Digital readiness assessment of italian smes: acase-study research. *Journal of Manufacturing Technology Management* 31(5), 1045–1083 (2020).
- [35] Bumann, J., Peter, M.: Action fields of digital transformation-a review and comparative analysis of digital transformation maturity models and frameworks. *Digitalisierung und andere Innovationsformen im Management* 2, 13-40 (2019)
- [36] Yezhebey, A., Sengirova, V., Igali, D., Abdallah, Y.O., Shehab, E.: Digital maturit and readiness model for kazakhstan smes. In: 2021 IEEE International Conference on Smart Information Systems and Technologies (SIST), pp. 1-6 (2021). IEEE
- [37] Amaral, A., Pe, cas, P.: A framework for assessing manufacturing smes industry 4.0 maturity. *Applied Sciences* 11(13), 6127 (2021)
- [38] Janssen, Z.-v.E.A.M.G. Van de Walle Cunningham: Benchmarking th digital maturity of the Dutch Health Insurance Companies. *://repository.tudelft.nl/islandora/object/uuid(2021)*
- [39] Mittal, S., Romero, D., Wuest, T.: Towards a smart manufacturing maturity model for smes (sm 3 e). In: *Advances in Production Management Systems. Smart Manufacturing for Industry 4.0: IFIP WG 5.7 International Conference, APMS 2018, Seoul, Korea, August 26-30, 2018, Proceedings, Part II*, pp. 155-163 (2018).Springer
- [40] Korachi, Z., Bounabat, B.: Data driven maturity model for assessing smart cities. In: *Proceedings of the 2nd International Conference on Smart Digital Environment*, pp. 140-147 (2018)
- [41] Rossmann, A.: *Digital maturity: Conceptualization and measurement model* (2018)
- [42] Asadamraji, E., Rajabzadeh GHatari, A., Shoar, M.: A maturity model for digital transformation in transportation activities. *International Journal of Transportation Engineering* 9(1), 415-438 (2021)
- [43] Berger, S., Bitzer, M., H'ackel, B., Voit, C.: Approaching digital transformationdevelopment of a multi-dimensional maturity model (2020)
- [44] [Al Hanaei, E.H., Rashid, A.: Df-c2m2: A capability maturity model for digital forensics organisations. In: 2014 IEEE Security and PrivacyWorkshops, pp. 57-60(2014). IEEE
- [45] Battista, C., Schiraldi, M.M.: The logistic maturity model: Application to a fashion company. *International Journal of Engineering Business Management* N5(God'i'ste 2013), 5–29 (2013)
- [46] Gill, M., VanBoskirk, S.: *The digital maturity model 4.0. Benchmarks: digital transformation playbook* (2016)
- [47] Schumacher, A., Erol, S., Sihh, W.: A maturity model for assessing industry 4.0 readiness and maturity of manufacturing enterprises. *Procedia Cirp* 52, 161-166 (2016)
- [48] Schumacher, A., Nemeth, T., Sihh, W.: Roadmapping towards industrial digitalization based on an industry 4.0 maturity model for manufacturing enterprises. *Procedia Cirp* 79, 409-414 (2019)
- [49] Ifenthaler, D., Egloffstein, M.: Development and implementation of a maturity model of digital transformation. *TechTrends* 64(2), 302-309 (2020)
- [50] Issa, A., Hatiboglu, B., Bildstein, A., Bauernhansl, T.: Industrie 4.0 roadmap: Framework for digital transformation based on the concepts of capability maturity and alignment. *Procedia Cirp* 72, 973-978 (2018)
- [51] [Langlo, J.A.-A., Sorskot, B., et al.: A suggested framework for measuring digital maturity in construction projects in norway. Master' s thesis, NTNU (2019).
- [52] Ilin, I., Borremans, A., Levina, A., Esser, M.: Digital transformation maturity model. In: *Digital Transformation and the World Economy: Critical Factors and Sector-Focused Mathematical Models*, pp. 221-235. Springer, ??? (2022)
- [53] R'usmann, M., Lorenz, M., Gerbert, P., Waldner, M., Justus, J., Engel, P., Harnisch, M.: Industry 4.0: The future of productivity and growth in manufacturing industries. *Boston consulting group* 9(1), 54-89 (2015)
- [54] Klisenko, O., Serral Asensio, E.: Towards a maturity model for iot adoption by b2c companies. *Applied Sciences* 12(3), 982 (2022).
- [55] Kusters, A.: Relating digitization, digitalization and digital transformation: a maturity model and roadmap for dutch logistics companies. B.S. thesis, University of Twente (2022)
- [56] Erdal, B., 'Ihtiyar, B., Mıstko'glu, E.T., G'ul, S., Temur, G.T.: Digital maturityya sssessment model development for health sector. In: *Digitizing Production Systems:Selected Papers from ISPR2021, October 07-09, 2021 Online, Turkey*, pp. 131–147 (2022). Springer



# Machine Learning based Predictive Modelling of Cybersecurity Threats Utilising Behavioural Data

## Cybersecurity Threat Predictive Modelling

Ting Tin Tin<sup>1</sup>, Khiew Jie Xin<sup>2</sup>, Ali Aitizaz<sup>3</sup>, Lee Kuok Tiung<sup>4</sup>, Teoh Chong Keat<sup>5</sup>, Hasan Sarwar<sup>6</sup>

Faculty of Data Science and Information Technology, INTI International University, Negeri Sembilan, Malaysia<sup>1</sup>  
Faculty of Computing and Information Technology, Tunku Abdul Rahman University of Management and Technology,  
Kuala Lumpur, Malaysia<sup>2</sup>

School of IT, UNITAR International University, Petaling Jaya, Malaysia<sup>3</sup>

Faculty of Social Science and Humanities, Universiti Malaysia Sabah<sup>4</sup>

DigiPen Institute of Technology Singapore<sup>5</sup>

Department of Computer Science and Engineering, United International University, Bangladesh<sup>6</sup>

**Abstract**—With the rapid advancement of technology in Malaysia, the number of cybercrimes is also increasing. To stop the increase in cybercrimes, everyone, including normal citizens, needs to know how secure they are while using digital appliances. A system is developed to predict the risk of users based on their behaviour when they are online using real-life behavioural data obtained from a private university's 207 undergraduates. Five supervised machine learning methods are being tested which are: Regression Logistics, K-Nearest Neighbour (KNN), Decision Tree (DT), Support Vector Machine (SVM), and Naïve Bayesian Classifier with the aid of a tool, RapidMiner. The algorithms are used to construct, test, and validate three categories of cybercrime threat (Malware, Social Engineering, and Password Attack) predictive models. It was found that KNN model produces the highest accuracy and lowest classification error for all three categories of cybercrime threat. This system is believed to be crucial in alerting users with details of whether the consumer behaviour risk is high or low and what further actions can be taken to increase awareness. This system aims to prevent the rise in cybercrimes by providing a prediction of their risk levels in cybersecurity to encourage them to be more proactive in cybersecurity.

**Keywords**—Cybersecurity threat; cybersecurity risk; predictive modeling; undergraduates; cybercrime

### I. INTRODUCTION

Malaysia has entered the digital age, with online meetings and classes or cashless payments becoming more popular [1]. However, as the number of digital users has increased over the years, it may also lead to a surge in cybercrimes. Although most Malaysians have a good level of awareness of cyber threats and risks, only a few who act against it due to a low understanding in cybersecurity and the severity of cyber threats and attacks [2] [3]. This high number of cyberattacks has been estimated to cost the global economy USD 1 trillion in 2020, that is, 50% more than in the previous year [4]. According to researchers, the increase in cybercrimes is also happening in Malaysia [5] [6]. Malaysia's cybersecurity is currently slow to catch up with the pace of advancement, and people lack of knowledge in cybersecurity due to the consequences and impacts of Malaysia's organisation, in the private or public

sector [5]. Cyberattacks go beyond the loss of money and reputation but remain a failure in finding a global systematic way to confront [7]. With numerous reports claiming that there is an increase in cybercrimes that are not only targeting important organisations and government but also normal citizens [7][8][9], there are various studies to warn digital users the don'ts and dos without certainly proclaiming how much precaution is needed to be considered safe in cyberspace.

The rise of cybercrimes in Malaysia has caused a lot of damage not only in terms of financial and reputation. However, as a normal citizen without any background knowledge in cybersecurity, it could be difficult for him to know and keep up to date with the latest cybersecurity news and may not even know where to start. One would need to read and listen to stories of victims of cybercrimes and learn from their mistakes to know the risks, but this is not enough because the sources of stories are limited as they were usually from the same social circle. This method of learning may be inaccurate and insufficient, as technology is advancing rapidly and may not be up to date with the latest cybersecurity methods. This concludes that there are no concrete means to prove one's knowledge of cyber risks in the current cybersecurity measures.

There are companies that offer cybersecurity services for companies to predict cyber threats and attacks using artificial intelligence (AI) and machine learning. Having that said, there is no need for normal citizens to hire a company just to know their risk levels in cybersecurity. Thus, the competition is scoped down to simple websites asking visitors a sample of questions to predict their awareness, as it is more non-tech-savvy friendly. These websites, however, do not have official databases and are opinion-based; no research is done in the prediction of results, but rather in a pop quiz-like structure. In general, these websites do not describe risks based on varied behaviours.

Furthermore, very little research was done among Malaysians and some existing research was outdated. Therefore, this study aims to fill this gap by helping researchers with cybersecurity prediction based on user

behaviour. Hence, the system would predict the user's risks based on real-life data sets and can give users an idea of which aspect of cyber risk is greater rather than only scores.

Predictive modelling of cybersecurity threats predicts the risks of a user while using a digital device such as a mobile phone, laptop and personal computer by using machine learning algorithms tested and validated by user behaviour data acquired from undergraduates in Malaysia. Therefore, the objective of this study is to identify the factors that affect users' security awareness (in terms of malware attacks, social engineering, and password attacks); build a predictive model of cybersecurity threats for undergraduates using the Internet in Malaysia; and develop a website to implement the predictive model.

This project has two main parts, data modelling and website deployment. Python is used to programme the machine learning part of the system, coded using Jupyter, with the dataset, while the website serves as a medium for users to predict their risks, who can access the Python files to make predictions. The results will then be displayed on the website. Web pages are structured using HyperText Markup Language(HTML) and the layout is formatted using Cascading Style Sheet(CSS) with JavaScript to give the web pages a final touch to make it more appealing in an integrated development environment (IDE), Visual Studio Code (VSC).

Young adults have a low understanding of the basics of cybersecurity as they are unfamiliar with common cyber threats [3]. The findings suggest that exposing cybersecurity knowledge at a young age can ensure healthy habits online, reducing the chance of cyber attacks and threats [2]. As gender does not affect prediction results because the prediction is entirely based on behaviours, this expands the system's target of the system to both genders.

This paper is constructed in four sections: Section II about literature review of current research in cybercrime prediction; Section III describes the methodology used in this present study; Section IV presents the study result and discussion; and Section V concludes the present study with limitation and future works.

## II. LITERATURE REVIEW

The digital economy dominates Malaysia in business transactions, as more than 40% of these transactions are made digitally. Research found that the future of Malaysia will depend on the digital economy; therefore, the digital space in Malaysia needs to be trusted to allow parties including enterprises, customers, public sectors and individuals to have a reliable digital space [1][10]. According to the Malaysia Computer Emergency Team, there is a visible increase in cyberattacks from January 2022 to July 2022 [11]. One of the recommendations to reduce the risk of cyber threats and attacks is the need to increase public awareness of risks, threats, and

vulnerabilities in cyberspace. Therefore, increasing awareness of cyber threats is one of the objectives of the system and is achieved by providing a prediction of user risk in cyberspace using predictive modelling with machine learning techniques.

Since there are various types of cyber threats, three threats, namely malware attack, social engineering, and password attacks, are selected due to the high likelihood of these threats against individuals. Other threats include advanced persistent threats (APT), where attackers gain unauthorised access to a network and try to become a part of the network to prevent detection for an extensive period of time, and Man-in-the-middle attack (MitM), where attacker intercepts users when they are remotely accessing a system over the Internet. APT requires attackers to possess a high knowledge of the victim, and therefore these attacks are usually launched towards nation states, large organisations, companies, or very important people. [12] As the target of this project is young undergraduate Internet users, they are less likely to connect their device remotely online, making MitM not in scope.

There are no equivalent research studies to the proposed project, but similar studies have been found that predict the cyber risk of software [13] [14]. Both projects use machine learning techniques to identify weak points or vulnerabilities in the system and the risk that the software becomes infected or corrupted at a certain time. Zhang et al. [13] built the predictive model with data from the National Vulnerability Database (NVD), which is a public data source for reported software vulnerabilities. They tested the data with various approaches for predictive modelling to find the best techniques for their prediction model, as their study results show that the current approach is not accurate except for a few vendors. Bilge et al. [14] on the other hand, got their data from 18 enterprises for a year, which contains information about binaries appearing on machines with fully and semi-supervised machine learning. Semi-supervised machine learning is a technique that uses machine learning machine learning that uses both supervised, where labelled data is used, and unsupervised, where unlabelled data are used [15].

Other work closely related to cyber risk prediction is cyberattack predictions and cyberattack detection. The prediction is usually overlooked by the research community opposed to cyberattack detection. Ben Fredj explored the prediction of cyberattacks using a deep learning approach [16]. It is a subgroup of the machine learning approach in which multiple layers of neural networks are used to build the model [15], which simulates how neuronal nerves work in a human brain. In addition to that, there is a study that surveyed not only machine learning approaches, but also data mining approaches in terms of prediction and prediction methods used in cybersecurity [17]. Regarding cyberattack detection [18][19][20], most studies use Deep Learning (DL) to model their data to detect which attacks will occur in given situations and the rate of these attacks (Table I).

TABLE I. SUMMARY OF RELATED WORKS

Study	Algorithm	Features/Factors	Reference
Software cyber risk prediction	Supervised Machine Learning	Identifying software vulnerabilities	Zhang et al., 2015; Bilge et al., 2017
Cyberattacks prediction	Deep Learning, Data Mining	Predicting cyberattacks	Ben Fredj et al., 2020; Husák et al., 2018
Cyberattacks detection	Deep Learning	Detecting cyberattacks	Berman et al., 2019; Moustafa et al., 2019; Aldweesh et al., 2020

### A. Malware

Malware means malicious software which refers to any software that intrudes on a system developed by cyber-attackers. This software can penetrate the device of a user ranging from viewing to modifying private data, such as user's personal photos, operating systems, and other data that the attacker can find on the victim's device [21]. Malware types include, but are not limited to, viruses, spyware, backdoor, and keyloggers, each with different threats and dangers. Viruses can attach themselves, using macros, to Microsoft Office software such as Words. Therefore, it infects the victim's computer when it is opened or viewed. Students will use Words frequently for various reasons such as completing assignments or recording notes, which pose a high possibility of becoming a victim. Other great possibilities include downloading free software online to avoid purchasing.

Malware is a programme that is inserted into a system with the intention of compromising the confidentiality, integrity, or availability of the victim's data, applications, or operating system, or otherwise annoying or disrupting the victim. Therefore, measuring risk in malware infection can be simplified to the ability to prevent malware from entering the system and the ability to mitigate threats if prevention fails. First, the ability to prevent malware can be measured by how many techniques the user knows about how a malware can enter a system and the depth of understanding of these techniques (MW1). Second, the ability to mitigate threats can be measured by how quickly the user can detect that malware has already entered the system, identified the source of the malware, and remove malware and its techniques (MW2) [22] [23]. Therefore, the system should collect the user response for the following regarding user's behaviour to avoid different malware threats:

- M1. Is antivirus software, firewall, and anti-spyware available on the user computers? (MW1, MW2)
- M2. What is the user's confidence level of antivirus software in their computers? (MW1, MW2)
- M3. How inclined is the user to download materials from unsecure sites? (MW1)
- M4. How inclined is the user to download freeware on the Internet? (MW1)
- M5. How inclined is the user to scan removable drives before using them on computers? (MW1)
- M6. How inclined is the user to apply security patches as soon as possible? (MW2)
- M7. Is the user able to sense that something is wrong if the computer runs oddly slow? (MW2)

### B. Social Engineering

The art of persuading people to breach information systems is known as social engineering. Instead of launching technical assaults on systems, social engineers use influence and persuasion to persuade people with access to information to reveal secret information or even carry out hostile actions. Most successful attacks on systems are rarely required to find technical vulnerabilities; hacking the human is usually sufficient [24]. Social engineering is the most successful when combined with other methods, such as phishing [25]. Phishing is the act of sending links that link victims to their website that do what cybercriminal programmes to do. For example, attackers send links decorated with official names and formatting to make them appear to come from a legitimate source to play mind tricks and get victims to click on the link. In addition to sending links, attackers can act as an advertiser trying to advertise a product and ask the victim to scan a QR code (quick response) that links to their malicious attack. Both situations are likely to occur amongst anyone.

Social engineering is a method of tricking victims to help compromise their own system. Therefore, user measurement of the risks in social engineering attacks can be simplified into the level of understanding of social engineering techniques and the ability to respond to these techniques correctly. First, how many social engineering techniques can the user know that can be used to measure the level of understanding (SE1). Second, whether the user knows how to respond to these techniques can be used to measure the ability to respond (SE2) [23] [26]. Therefore, the system should collect the user response for the following:

- S1. Is the user interested in learning social engineering issues? (SE1, SE2)
- S2. Does the user establish a trusted relationship with strangers on-line? (SE1, SE2)
- S3. How inclined is the user to click hyperlinks in email messages? (SE1, SE2)
- S4. How inclined is the user to check the authorisation of the interlocutor? (SE1)
- S5. How inclined is the user to check URL spellings? (SE1)
- S6. Does the user trust any benefit winning emails, calls, or SMS? (SE1)
- S7. Does the user trust in any information online? (SE1)
- S8. Is the user aware of the latest scam and phishing techniques? (SE1)
- S9. Does the user feel intimidated by questions by any interlocutor? (SE2)
- S10. How inclined is the user to provide details to authorities? (SE2)



S11. How inclined is the user to respond to calls, SMS, or email from strangers? (SE2)

### C. Password Attack

Password attacks occur when attackers attempt to gain access to a victim's system using the victim's password. This attack is different from the above two threats, as this threat attacks through the 'front door' rather than in secret or stealthily by guessing and trying repetitively until it is correct. User passwords are easy to guess, since they are related to the victim or the password is an actual word or phrase [27], which can be easily obtained using social engineering techniques. As Malaysia is moving toward a digital era, account creation can be common and logging in or signing up requires a password. Other techniques of password attacks include, but are not limited to, brute force, where the attackers try every possible password combination, or dictionary attack, where attacks steal the encrypted data during transmission containing the victim's password and decrypt it using their encryption library.

Password attack is a method to legitimately enter the victim's system through victim passwords. Therefore, to measure the risks of users in password attacks, it can be measured by how securely users keep their passwords and complexity [23] [28]. Therefore, the system should collect the user response for the following:

- P1. Does the user's password follow a keyboard pattern?
- P2. Does the user share passwords with other people?
- P3. Does the user create different passwords for different applications?
- P4. Is the user's password consisting of lowercase, uppercase, numbers, special characters?
- P5. Is the user password longer than 8 characters?
- P6. Is the user's password created based on personal/information?
- P7. Does the user change the password?
- P8. Does the user use the 'Recall password' option?
- P9. Does the user write the password?
- P10. Does the user use 'hint' to recover forgotten passwords?
- P11. Does the user check for a padlock symbol on browsers?

### D. Conceptual Framework for Measuring Ability to Avoid Cyberattacks

Based on the literature review, Fig. 1 summarises the measurement criteria of ability to avoid cyberattacks of three categories of threats – malware, social engineering, and password attack.

### E. Web Projects

Moving from research-based projects to web projects, three web services, namely ProProfs, W3Schools, and the Federal Trade Commission (FTC), are being compared as follows. This predictive modelling is built for young people in Malaysia, which is the scope that is not covered by these three websites.

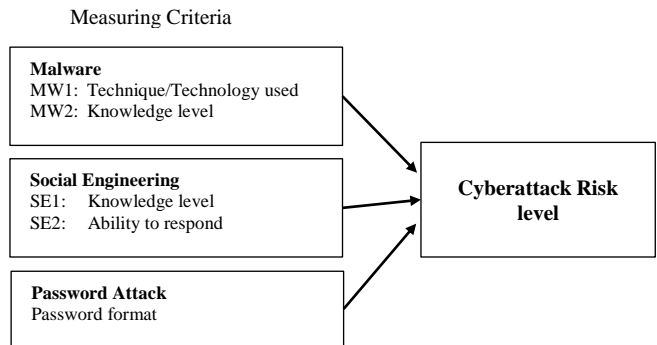


Fig. 1. Conceptual framework of measuring criteria for cyberattack risk level.

First, ProProfs is a website that allows any user to create quizzes and post them online on the ProProfs website itself (Fig. 2). Therefore, this website has a variety of quizzes from different domains, which, of course, includes cybersecurity. However, most of the questions of these quizzes are focused on cybersecurity as a course instead of a test for user risks on-line. The questions asked are technical and not suitable for general users who do not consider cybersecurity as their focus. On the ProProfs website, a quiz is found that tests for users' cyber health and security, but it seems to have the same results for all responses entered. Since it is available to everyone, most of these quizzes do not have concrete backing of data to support the claims of the results.

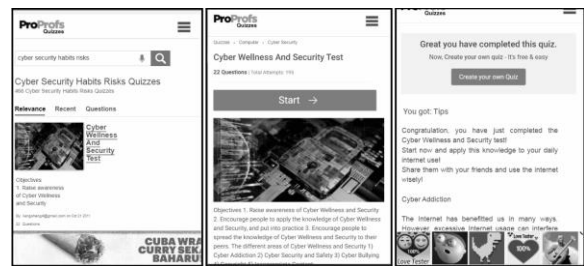


Fig. 2. Screenshots of the ProProfs website.

Next, W3Schools is a free educational website to learn Python coding (Fig. 3). However, this website is controlled by two entities namely Refsnes Data and W3schools Network instead of a central point for everyone to submit their viewpoints. As mentioned, this website is built for educational purposes and the cybersecurity quiz is one of the many quizzes found, which is also for people who want to make a revision of cybersecurity courses. Therefore, it does not inform users about the cyber risks that could occur to users.

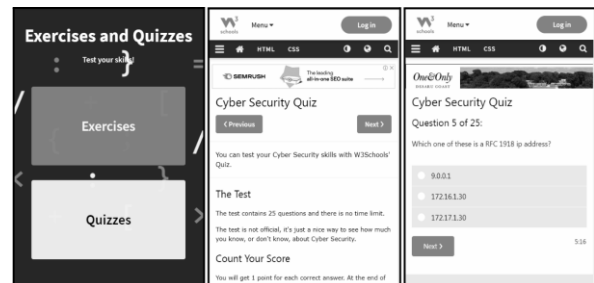


Fig. 3. Screenshots of W3Schools' website.

The Federal Trade Commission is an official website of the United States (USA) government that is built to protect American consumers (Fig. 4). It contains cybersecurity quizzes for small businesses to help guide them. The topics in the cybersecurity quizzes are the basics of cybersecurity, physical security, ransomware, phishing, vendor security, and secure remote access. In addition to quizzes, it also provides other means of guidance, such as but not limited to downloadable publications and videos of cybersecurity, which are all accessible in the additional resource's subsection of the page. Table II summarises the three websites in terms of owner, target users, location, and content(s).

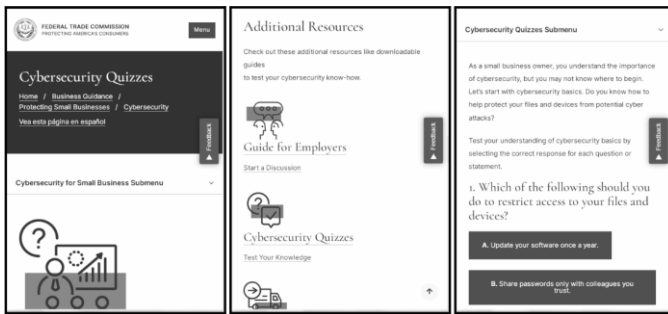


Fig. 4. Screenshots of FTC's website.

TABLE II. COMPARISON OF 3 WEB SERVICES WITH THE PROPOSED PROJECT

	ProProfs	W3Schools	FTC
<b>Owner(s) of Content</b>	Anyone	Refsnes Data and W3schools Network	US Government
<b>Target Users</b>	Not specific	Learner	Small business
<b>Location</b>	Not specific	Not specific	US
<b>Content(s)</b>	Quizzes	Quizzes Guidance	Quizzes Guidance

### III. METHODOLOGY

Questionnaire items are modified to avoid multiple-choice types of questions. Its purpose is to overcome the limitations of multiple-choice questions, which are the excessive words that make users feel more like an exam and will try to give a 'correct answer' instead of their genuine online behaviour. The data entry designs are shown in Table III.

In this study a private university in Malaysia in the age group of 15 to 30 years constituted the population. The sampling plan implemented in this investigation is the simple random sampling method (SRS). A total of 207 undergraduates participated in the study.

TABLE III. MODIFIED QUESTIONS

ID	Question	Response Type	Measured Questions
<b>Malware</b>			
L1	Is your device's operating system (OS) up-to-date?	5 likert scale	M6, M2
L2	Do you scan removable drives?		M5
L3	Do you download freeware online?		M3, M4
L4	Do you feel something is wrong if your device is running slow?		M7, M2, M4
L5	Is your device protected by any cybersecurity measures?		M1, M4
<b>Social Engineering</b>			
E1	Are you interested in learning about social engineering issues?	5 likert scale	S1,S8
E2	Do you establish a trusted relationship with strangers online?	5 likert scale	S2, S11, S8
E3	Do you click on links in emails?	5 likert scale	S3, S7, S8
E4	Do you check the authorisation of the authorities?	5 likert scale	S4, S7, S8
E5	Which link is the right URL to the Google website? www.google.com; google.com; https://google.com; g00gle.com; http://google.com		S5, S8
E6	Do you feel intimidated by questions from any authority?	5 likert scale	S9, S11
E7	Do you provide details to the authorities?	5 likert scale	S10, S8
<b>Password Attacks</b>			
A1	Create a password that you will use.	Open ended	P1, P4, P5
A2	Is the password created based on personal/ information?	5 likert scale	P6
A3	Do you change your password?		P7
A4	Do you use password management features?		P8, P10
A5	Write the password?		P9
A6	Do you share passwords?		P2
A7	Do you check for a padlock symbol on browsers?		P11
A8	Do you create different passwords for different applications?		P3

Participants responded to the questionnaire based on a 5-point Likert scale, which divides into 5 categories (strongly agree, agree, neither agree, disagree, strongly disagree). The questions are then analysed to determine whether they are good or bad practises. For questions classified under good practises, the mark is allocated accordingly based on the options (“Strongly agree”-5, “Agree” - 4, “Neutral” - 3, “disagree”- 2, “Strongly disagree” - 1) while for questions classified under bad practises, the mark allocated for each option is the opposite of good practises (“Strongly agree”-1, “Agree” - 2, “Neutral” - 3, “disagree”- 4, “Strongly Disagree” - 5). The responses to every question may vary; however, they generally have the same meaning. The scores for each question for each category are summed up as a total score. Thus, the highest scores attainable on the questionnaire for Malware, Social Engineering, and Password Attack are 25, 35, 40 respectively (best cybersecurity practises implemented), and the lowest scores are 5, 7, 8 respectively (worst cybersecurity practises implemented). Data are then statistically transformed into maximum scores of 35, 55, 55 and lowest 7, 11, 11 respectively.

Questions are either the main question itself, thus not needing to be processed, or are paired with other questions. Questions that are paired with others are calculated using the mean of all questions related to it, except questions A1 and A4. For example, questions L1 and L4 also have value for question M2. Therefore, the value of M2 to be given to the model is the mean value of L1 and L4.

For question A1, the input text will be used to measure 3 parameters.

1) For input text that follows a keyboard pattern, will be marked as low score, while a text that does not will be marked as a high score.

2) The length of the text will determine the score for P4. To achieve the best score (5), users must have an input text of more than 16 characters, while less than 4 characters will be marked as low score(1).

3) The number of character types will determine the score for P5. Text input will be marked as the best score (5) if it contains all types of character (lower case, upper case, numbers, special characters) and the lowest score(one) if it only contains one type of character.

Regarding questions A4, P8 and P10, they point to similar features that most applications provide, which are ‘remember password’ and ‘forget password’. Thus, both scores will be equal. The model will receive the user's behaviour in cyberspace as input to determine its awareness and then predict the user's risk of cyber threats (Fig. 5).

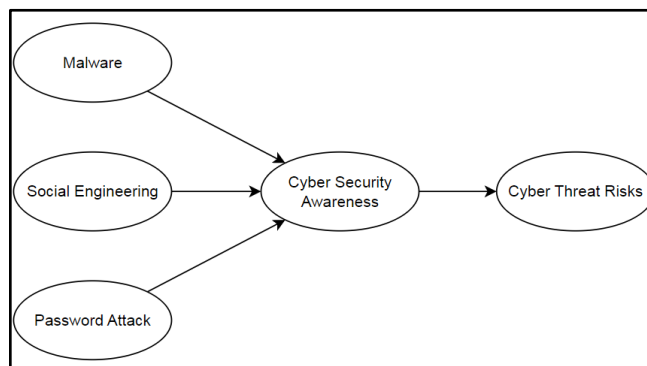


Fig. 5. Predictive Modelling components of Cyber Threats.

#### IV. RESULT AND DISCUSSION

##### A. Model Performance

Based on several related works that have been studied, a supervised machine learning method has been selected for the predictive model (Fig. 6). Five supervised machine learning methods are being tested, Regression Logistics, K-Nearest Neighbour (KNN), Decision Tree (DT), Support Vector Machine (SVM), and Naïve Bayesian Classifier with the aid of a tool, RapidMiner. K-fold cross-validation, where the dataset is divided into 5 groups with each group being the test data set after training the machine with other groups, is used to assess every method above. Of the above five, KNN is selected as the machine learning methodology, as it has the highest accuracy among the other methods (Table IV). KNN is an algorithm that calculates the distance between the new data point and the nearest available data point, where k is a positive integer. The new point is then classified according to which class has the most data points closest to the new data point. The contingency table or confusion matrix is used to help display the accuracy of all the above-mentioned methods. The accuracy is calculated with formula 1 and simplified with formula 2 into percentage (%).

Formula 1:

$$\frac{TruePositives}{TotalPredictedYes} + \frac{TrueNegatives}{TotalPredictedNo} = Accuracy$$

Formula 2:

$$TruePositives + TrueNegatives = Accuracy$$

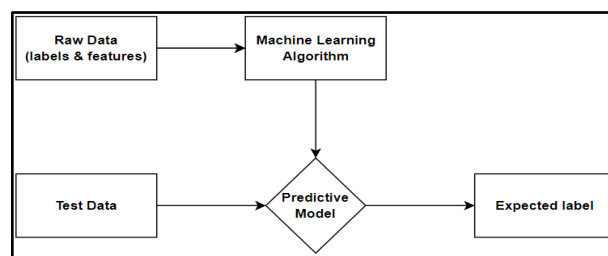


Fig. 6. Supervised machine learning model.

TABLE IV. SUMMARY OF 5 MODELS FOR EACH CATEGORY OF CYBERATTACKS

Model	Malware		Social Engineering		Password Attack		Average Accuracy (%)
	Accuracy (%)	Classification Error (%)	Accuracy (%)	Classification Error (%)	Accuracy (%)	Classification Error (%)	
Naïve Bayesian Classifier	79.5	20.5	89.8	10.2	91.4	8.6	86.9
Regression Logistics	79.5	20.5	88.2	11.8	91.4	8.6	86.4
KNN	92.9	7.1	93.8	6.2	97.6	2.4	94.8
DT	83.0	17	91.5	8.5	79.5	20.5	84.7
SVM	83.0	17	91.5	8.5	81.2	18.8	85.2

**B. Model Fit**

Python has been selected as the programming language for the machine learning part of the system. Python is selected because it has built-in libraries and frameworks suitable for data science. The libraries used for this project are pandas, NumPy, and Scikit-learn. Pandas library is used to read data tabulated in excel sheets, NumPy is used to process the data into machine learning parameters for the model to train, and Scikit-learn is used to implement machine learning models. A built-in Python module, pickle, is used to save the model as a non-readable binary file to be placed in the server and accessed by the webpage. Two parameters are needed to train the model, the first being the data to test, while the second being the K values.

To get the first parameter, the data is loaded into memory with pandas extracting data from excel sheets. Numpy is then used to convert the data into arrays. These arrays are then divided into training and test data with a ratio of 5: 1 (80% training, 20% testing). Training data will be used to fit the model while test data are used to measure the accuracy of the model.

The next parameter is to find the best K-value for the model. For this, another two-array list is created, namely a set of K values, from 3 to 30, and an empty list to store the results. The model is trained 28 times, and its result score is stored in the empty array list. The least K value with the highest accuracy is then selected as the K value. A lower value of K means that the classification is close to the original value and will not include further away data points, thus increasing precision. K values are evaluated as shown in Fig. 7, 8, and 9. The K values are 4, 5 and 3 for malware attack, social engineering, and password attack model, respectively. After splitting the data set and finding the optimal K values, the model is ready to be saved as a binary file using pickle.

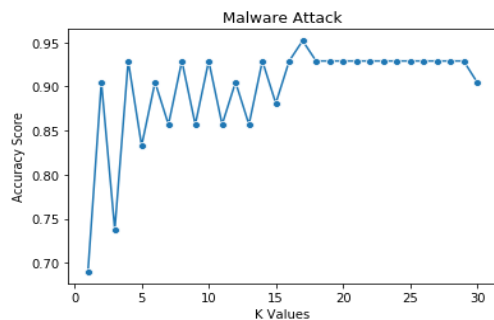


Fig. 7. Accuracy score of K values for malware attacks.

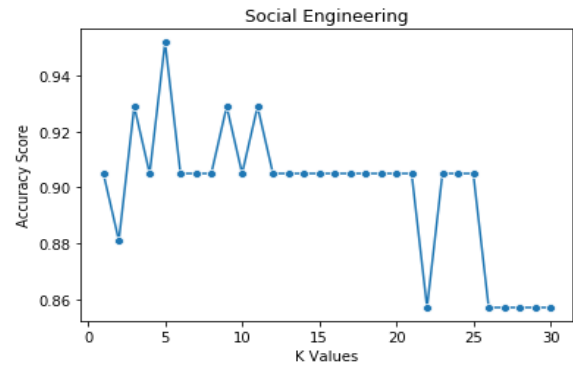


Fig. 8. Accuracy score of K values for social engineering.

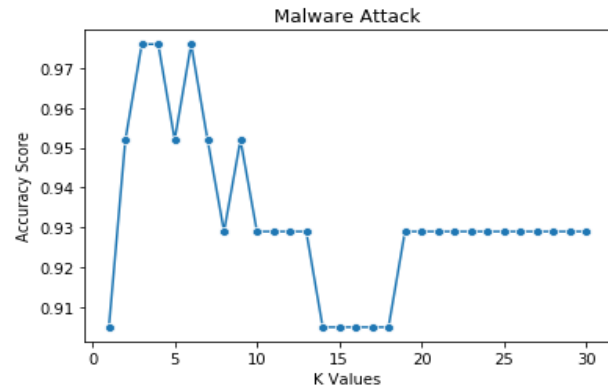


Fig. 9. Accuracy score of K values for malware attacks.

**C. Validity and Reliabilty Consideration**

Prior to this actual study, the instrument (questionnaire) was pilot tested with a group of 30 students from the same research site. The researchers ensure that the participants for the pilot test do not participate in the actual study. Data collected from the pilot test were measured for reliability using the Cronbach alpha reliability coefficient, formula 3.

Formula 3:

$$a = \frac{k}{k - 1} \left( 1 - \frac{\sum V_i}{V_t} \right)$$

Where a is the reliability coefficient, k is the number of questions,  $V_i$  is the variance of the responses of each question, and  $V_t$  is the variance of the total score of each respondent. The reliability measurement of the questionnaire is shown in Table V where all categories show good reliability.

TABLE V. RELIABILITY MEASUREMENT

Cyber Security Categories	Number of Items	Reliability Coefficient
Malware	5	84.3%
Social Engineering	7	81.0%
Password Attack	8	80.2%

D. Demographic Variable Effect

This study is specifically aimed at the demographics of users (age, geography). The variable (gender) is not used in the prediction; however, this variable is still being tested by independent samples T-Test to compare the means of two independent groups (male and female) to prove that this variable does not affect the accuracy of the prediction (Table VI). Data are statistically analysed using the SPSS programme. A total of 207 Malaysians participated in the study, of whom 98 of the respondents are male and 109 of the respondents are female.

TABLE VI. VALUES OF THE T-TEST FOR DIFFERENCES IN THE LEVEL OF CYBER SECURITY BEHAVIOUR BY GENDER IN THE ASPECT OF MALWARE, SOCIAL ENGINEERING, AND PASSWORD ATTACK

	Group statistics		T-Test	
	Mean	Std Deviation	t-value	Sig.
<b>Malware</b>				
Male	33.60	5.49	0.05	0.963
Female	33.57	4.76		
<b>Social engineering</b>				
Male	34.28	5.27	-1.02	0.306
Female	34.99	4.76		
<b>Password Attack</b>				
Male	33.46	5.79	0.26	0.796
Female	33.25	5.96		

Based on the three tables above, there are no significant gender differences in the level of cybersecurity behaviour in all three aspects (malware, social engineering, password attack); thus, gender does not affect the accuracy of the prediction of the model.

E. Comparing Proposed Website with Existing Websites

1) *Similar existing website:* Three previously mentioned websites are studied for their functionalities to choose those that are applicable for this project. All of them follow the same flow, that is, to introduce what and how important cybersecurity is and a navigation panel or menu which links to other functionalities.

The Proprofs website allows users to view the list of questions of the selected quiz, contact the author of the quiz, take the quiz, edit the settings of the webpage, to search for other quizzes from any domain, to share the selected quiz, in embedding the quiz to another website. The Proprofs website also allows users to create quizzes with the precondition of having an account with Proprofs, thus needing users to log in prior to creating quizzes [29].

The W3schools website allows users to view an introduction of cybersecurity, search for other services that w3school provides, log into a w3school account, take the quiz,

quick link to access other tutorials, change the theme of the website, translate the website to another language. The W3schools website also allows users to subscribe to their services under the condition that the user has an account with w3schools, which requires that the users log in [30].

The FTC website allows users to translate the website into another language, report fraud, sign up for FTC newsletters, search for other documents in a legal library, give feedback, view Introduction to Cybersecurity, print the website, take the risk assessment, and access to other services [31].

Based on the study above, all websites have similar design and functions; therefore, to be consistent with the existing websites, the Predictive Modelling of Cyber Security Threats website should display an introduction to cybersecurity and explanation of its importance (Fig. 10). The navigation menu should also be added to this website for easy navigation between other web pages, which includes displaying the information page and the methodology page. Users can also choose to share this website. Feedback from user functions should also be included so that this website can interact with users for future improvements. Lastly, the website should allow users to assess their cyber risks. Other functions such as searching, log-in or log-out, and printing are omitted in this website, as they serve no purpose for their functions on this website. For example, this website does not need a search function, as this project has only cyber risk prediction as its focus. The comparison is summarised in Table VII.



Fig. 10. Screenshot of the project website prediction result page.

TABLE VII. FEATURES OF THE ABOVE 3 WEBSITES AND PROPOSED SYSTEM

Features	Proprofs	w3schools	FTC website	Proposed website
Login / Logout	✓	✓		✓
Subscription	✓	✓		✓
Display cyber info	✓	✓	✓	✓
Quiz / Assessment	✓	✓	✓	✓
Share website	✓			✓
Guides		✓	✓	✓
Webpage translation			✓	✓
Feedback			✓	✓
Display methodology				✓
Machine learning				✓
FAQ				✓

## V. CONCLUSION

This project uses five machine learning algorithms (Regression Logistics, K-Nearest Neighbour (KNN), Decision Tree (DT), Support Vector Machine (SVM), and Naïve Bayesian Classifier) to predict the risk of cyber threats in the aspects of malware attack, social engineering, and password attacks among Internet users based on their online behaviour. During the development of this present study, it was also found that gender does not play a role in the perception of cybersecurity in Malaysia. KNN predictive model produced the highest accuracy and the lowest classification error. Therefore, KNN model is further improved using Python.

Given the absence of previous studies utilizing machine learning techniques for predicting users' cyberattack risk levels, this present study introduces a conceptual framework that includes measurement criteria for assessing risk levels. Most of the previous studies are predicting the cyberattacks of organisation websites or companies' networks instead of individual risk level. This study serves as guidance for future researchers to continue the study in other cyberattacks such as MitM. New behaviours can also be incorporated to investigate cyber risks. Furthermore, this present study only focused on a data set of young people, since all participants in this project were in the age group of 15 to 30. More efforts are needed in this domain, as predicting human behaviour is a complex task [10]. Techniques to detect potential cyberattacks are crucial to ensure a safe world of the Internet for global users.

## REFERENCES

- [1] Mat, B., Pero, S., Wahid, R., and Sule, B., 2019. Cybersecurity and the digital economy in Malaysia: trusted law for customer and enterprise protection. *International Journal of Innovative Technology and Exploring Engineering*, 8(3), pp.214-220.
- [2] Zulkifli, Z., Molok, N.N.A., Abd Rahim, N.H. and Talib, S., 2020. Cyber security awareness among secondary school students in Malaysia. *Journal of Information Systems and Digital Technologies*, 2(2), pp.28-41.
- [3] Fatokun, F.B., Hamid, S., Norman, A. and Fatokun, J.O., 2019. The impact of age, gender, and educational level on the cybersecurity behaviors of tertiary institution students: an empirical investigation on Malaysian universities. *Journal of Physics: Conference Series*, 1339(1), p. 012098. <https://doi.org/10.1088/1742-6596/1339/1/012098>
- [4] Cremer, F., Sheehan, B., Fortmann, M., Kia, A.N., Mullins, M., Murphy, F. and Materne, S., 2022. Cyber risk and cybersecurity: a systematic review of data availability. *The Geneva Papers on Risk and Insurance-Issues and Practice*, pp.1-39. <https://doi.org/10.1057/s41288-022-00266-6>
- [5] Teoh, C.S., Mahmood, A.K. and Dzazali, S., 2018. Cyber security challenges in organizations: a case study in Malaysia. 2018 4th International Conference on Computer and Information Sciences, pp. 1-6.
- [6] Abdullah, F., Mohamad, N.S. and Yunos, Z., 2018. Safeguarding Malaysia's cyberspace against cyber threats: contributions by cybersecurity Malaysia. *OIC-CERT Journal of Cyber Security*, 1(1), pp.22-31.
- [7] Singh, M.M., Frank, R. and Zainon, W.M.N.W., 2021. Cyber-criminology defense in pervasive environment: a study of cybercrimes in Malaysia. *Bulletin of Electrical Engineering and Informatics*, 10(3), pp.1658-1668.
- [8] Khan, S., Khan, N. and Tan, O., 2020. Efficiency of legal and regulatory framework in combating cybercrime in Malaysia. In *Understanding Digital Industry*, pp. 333-336. Routledge.
- [9] Isa, M.Y.B.M., Ibrahim, W.N.B.W. and Mohamed, Z., 2021. The relationship between financial literacy and public awareness on combating the threat of cybercrime in Malaysia. *The Journal of Industrial Distribution & Business*, 12(12), pp.1-10.
- [10] Sulaiman, N. S., Fauzi, M. A., Hussain, S., & Wider, W., 2022. Cybersecurity behavior among government employees: The role of protection motivation theory and responsibility in mitigating cyberattacks. *Information*, 13(9), 413. MDPI AG. <http://dx.doi.org/10.3390/info13090413>
- [11] MyCERT, 2022. MyCERT Incident Report 2022.
- [12] Cassetto, O., 2023. Cybersecurity threats: Types and challenges, Exabeam. Available at: <https://www.exabeam.com/information-security/cyber-security-threat/>
- [13] Zhang, S., Ou, X. and Caragea, D., 2015. Predicting cyber risks through national vulnerability database. *Information Security Journal: A Global Perspective*, 24(4-6), pp.194-206. <https://doi.org/10.1080/19393555.2015.1111961>
- [14] Bilge, L., Han, Y. and Dell'Amico, M., 2017. Riskteller: Predicting the risk of cyber incidents. In *Proceedings of the 2017 ACM SIGSAC conference on computer and communications security*, pp. 1299-1311. <https://doi.org/10.1145/3133956.3134022>
- [15] Zhou, Z.H., 2021. *Machine learning*. Springer Nature.
- [16] Ben Fredj, O., Mihoub, A., Krichen, M., Cheikhrouhou, O. and Derhab, A., 2020. CyberSecurity attack prediction: a deep learning approach. 13th International Conference on Security of Information and Networks, pp. 1-6. <https://doi.org/10.1145/3433174.3433614>
- [17] Husák, M., Komárková, J., Bou-Harb, E. and Čeleda, P., 2018. Survey of attack projection, prediction, and forecasting in cyber security. *IEEE Communications Surveys & Tutorials*, 21(1), pp.640-660. <https://doi.org/10.1109/COMST.2018.2871866>
- [18] Berman, D.S., Buczak, A.L., Chavis, J.S. and Corbett, C.L., 2019. A survey of deep learning methods for cyber security. *Information*, 10(4), p.122. <https://doi.org/10.3390/info10040122>
- [19] Moustafa, N., Hu, J. and Slay, J., 2019. A holistic review of network anomaly detection systems: A comprehensive survey. *Journal of Network and Computer Applications*, 128, pp.33-55. <https://doi.org/10.1016/j.jnca.2018.12.006>
- [20] Aldweesh, A., Derhab, A. and Emam, A.Z., 2020. Deep learning approaches for anomaly-based intrusion detection systems: A survey, taxonomy, and open issues. *Knowledge-Based Systems*, 189, p.105124.
- [21] Lutkevich, B., 2022. What is malware? definition, types, prevention - techtarget, Security. TechTarget. Available at: <https://www.techtarget.com/searchsecurity/definition/malware>
- [22] Martens, M., De Wolf, R. and De Marez, L., 2019. Investigating and comparing the predictors of the intention towards taking security measures against malware, scams and cybercrime in general. *Computers in Human Behavior*, 92, pp.139-150. <https://doi.org/10.1016/j.chb.2018.11.002>
- [23] Muniandy, L., Muniandy, B. and Samsudin, Z., 2017. Cyber security behaviour among higher education students in Malaysia. *J. Inf. Assur. Cyber Secur*, 2017, pp.1-13.
- [24] Mann, I., 2017. *Hacking the human: social engineering techniques and security countermeasures*. Routledge.
- [25] Abass, I.A.M., 2018. Social engineering threat and defense: a literature survey. *Journal of Information Security*, 9(04), p.257.
- [26] Albladi, S.M. and Weir, G.R., 2020. Predicting individuals' vulnerability to social engineering in social networks. *Cybersecurity*, 3(1), pp.1-19. <https://doi.org/10.1186/s42400-020-00047-5>
- [27] Tasevski, P. and Eurecom, F., 2015. Methodological approach to security awareness program. In *CyberSecurity for the Next Generation Conference*.
- [28] Ye, B., Guo, Y., Zhang, L. and Guo, X., 2019. An empirical study of mnemonic password creation tips. *Computers & Security*, 85, pp.41-50. <https://doi.org/10.1016/j.cose.2019.04.009>
- [29] ProProfs.com. Available at: <https://www.proprofs.com/>
- [30] W3Schools.com. Available at: <https://www.w3schools.com/>
- [31] Federal Trade Commission. Available at: <https://www.ftc.gov/>

# Enterprise Marketing Decision: Advertising Click Through Rate Prediction Based on Deep Neural Networks

Luyao Zhan

Faculty of Business and Accountancy, Henan Open University, Zhengzhou, 450008, China

**Abstract**—With the high-speed growth of modern information technology, online advertising, as a new form of advertising on the Internet, has begun to emerge, demonstrating enormous development potential. To improve the accurate estimation of advertising placement and improve the operational efficiency of the advertising placement system, an improved deep neural network model for forecasting advertising click through rate was studied and designed. Meanwhile, the values of the activation function and the parameter dropout are determined, and the prediction accuracy of the deep neural network model and the improved model is compared and analyzed. The experimental results show that the training time of the improved prediction model has been shortened by about 73.25%, resulting in a significant improvement in computational efficiency. When the number of iterations is 110, the logarithmic loss function value is 0.208, and the logarithmic loss function value of the improved model is 0.207, with an average loss reduction of 0.4%. In the area comparison under the receiver operating characteristic curve, the pre improved model was 0.7092, and the improved model was 0.7207. Meanwhile, compared to before the improvement, the prediction accuracy of the improved model increased by 1.6%. The data validates that the optimized model has high prediction precision and efficiency, and has certain application potential and commercial value in marketing.

**Keywords**—Click through rate prediction; deep learning; deep neural network; online advertising; marketing

## I. INTRODUCTION

Nowadays, the growth of internet companies has been inseparable from advertising marketing, and Click Through Rate (CTR) prediction remains a key issue in the advertising field. With the continuous improvement of internet commerce and search engines, online advertising has become one of the main ways for businesses to promote and market [1]. With the development of information technology such as the Internet and intelligent terminals, the scale of the domestic advertising industry market has been continuously expanding in the past few years. The consecutive expansion of the advertising industry has driven the sustained growth of internet advertising. In internet marketing, CTR is the ratio of the click numbers on a certain content on a website page to the quantity of times. It shows the level of attention paid to a certain content on a webpage and is taken to measure the attractiveness of advertisements [2]. Based on information such as user behavior attributes and advertising characteristics, a prediction model can be constructed using deep learning methods for advertising CTR prediction [3]. Deep learning

methods have achieved good results in speech and image recognition, and can also reduce manual repetition in the field of advertising CTR prediction, controlling the accuracy and efficiency of advertising production [4]. The advantage of Deep Learning Model (DLM) in automatically extracting higher-order features improves accuracy, but it lacks a certain degree of interpretability compared to manually extracted features [5]. Therefore, this study designed the advertising CTR prediction model of the Deep Neural Network based on Sampling (SDNN), and determined the activation function and the parameter dropout. This article also compares and analyzes the prediction accuracy of four models: Deep Neural Network (DNN), SDNN, DNN trained with dropout, and SDNN trained with dropout. The research aims to improve the operational efficiency of the advertising placement system, so that enterprise marketing can determine more accurate decisions. The research content mainly includes six sections. The second section is a review of the current research status of advertising CTR prediction and deep learning both domestically and internationally. The third section constructs an improved DNN advertising CTR prediction model. Under this, the first sub-section designs an advertising CTR model, and the second sub-section proposes an SDNN advertising CTR prediction model. Section four analyzes the results of the improved DNN advertising CTR prediction model. The fifth section is a discussion on improving the advertising CTR prediction model of DNN. Section six is the conclusion.

## II. RELATED WORKS

The high accuracy of CTR prediction can help advertisers and advertising platforms increase revenue and gain greater benefits. Many researchers have proposed various ideas and methods here. Cai et al. constructed a neural network and global attention mechanism model to precisely forecast the possibility of users clicking on advertisements, achieving interaction between low and high order nonlinear features, while promoting optimization of deep structures. This model has good predictive performance [6]. Xue and other professionals have constructed adaptive hash algorithms and DNN models to reduce redundant features in deep CTR prediction problems. It can automatically select practical features for high-order interaction. This model has good performance, low complexity, and requires less training time [7]. Liu and other scholars proposed an A-CTR-P that combines big data analysis to achieve mobile computing of advertising CTR logs, and used power-law distribution for log preprocessing and category feature extraction. This model has

good prediction accuracy [8]. Zhou's team proposed a model that combines advertising topic distribution network and recurrent neural network to control data transmission for e-commerce product advertising recommendation. This model has high accuracy while reducing computational complexity [9]. To improve advertising profits and promote user experience, Ghorbel et al. built the upper confidence limit and A-CTR-P of GA built on the LSTM network to improve the feature selection of micro targeting technology and optimize hyper-parameter. The accuracy of this method reaches 87%, the precision reaches 89%, and the recall rate reaches 92% [10]. Liu et al. designed a user preference network model for a recommendation system that combines attention for CTR prediction under video recommendation. This model has good predictive performance and solves the time series problem of user feedback information [11].

Deep learning has made good progress in the image and natural language processing, and is widely used in daily life. Hung's team proposed an early warning model that combines machine learning and deep learning algorithms to identify learners at risk in order to analyze student behavior performance. This model captured 59% of high-risk students, with an overall accuracy of 86.8% [12]. Ma et al. built transfer learning and deep learning models for the prediction of residual life transfer of batteries with different formulations to avoid the loss of battery information. It saved testing costs and ensures high temperature robustness, with high prediction accuracy, while optimizing batteries with different formulations [13]. To avoid employee turnover, professionals such as Ozmen EP have constructed a hybrid extended convolutional decision tree model with good classification accuracy based on convolutional neural networks and grid search optimization [14]. Deng and other scholars designed a topology optimization method combining geometric depth learning to elaborate the density distribution function for compliance and pressure constraints. It ensured the boundary smoothness and effectively reduces design variables and controls structural complexity, making it very practical [15]. Hu et al. raised a deep reinforcement learning method driven by curiosity to optimize intelligent and interconnected automotive power control systems to accelerate training speed and achieve a good balance of universality. Under this algorithm, the control behavior rate had been optimized by 50.43%, and the learning productivity had been improved by 74.29% [16]. To optimize the energy efficiency of cooperative spectrum sensing, He H's team has built a high-performance reinforcement learning and deep learning framework for graphical neural networks to promote the improvement of system energy efficiency and spectrum efficiency [17].

In summary, many professionals have conducted research on CTR prediction and deep learning, and applied them to various fields. However, there are still few research results on using deep learning to predict CTR, and this direction has strong potential application value for lifting the accuracy of advertising prediction.

### III. A-CTR-P BASED ON DNN

To rise the advertising prediction accuracy and the operational efficiency of the advertising delivery system,

A-CTR-P is designed to improve DNN, and the activation function and dropout values are determined.

#### A. Construction of A-CTR-P

In today's era, traffic monetization and product promotion depend on advertising. The precise recommendation of advertisements helps to improve user experience and the ability to monetize platform traffic. The key factor in achieving precise recommendations is CTR estimation. The accuracy of CTR estimation affects the decisions of advertisers and advertising platforms regarding advertising placement, thereby determining bids based on user click behavior. When advertising platforms calculate the bidding price of each advertiser, CTR estimation is the core link. The relationship between internet platform users and advertisers is Fig. 1.

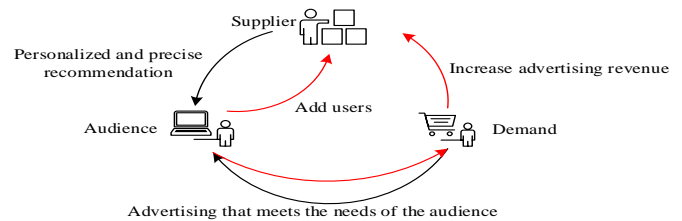


Fig. 1. Calculate the tripartite relationship of advertising.

In Fig. 1, internet platforms enhance user experience and increase user stickiness through personalized and precise recommendations. The user's click behavior on the pushed product is used as feedback to promote the optimization of the push product mechanism on various internet platforms. This can increase platform revenue while improving user experience. The data related to CTR prediction comes from the Criteo dataset. 0 means the user without clicking, and 1 indicates that the user clicked. There are 13 numerical features and 26 string features in the data [18]. Before building a prediction model, it is necessary to complete data processing and feature engineering. The data preprocessing process involves descriptive statistics of the overall data, as Fig. 2.

In Fig. 2, first of all, the data is processed separately according to the type in the database, and the string data is characteristic processed, that is, One-hot coding. The data that has been uniquely hot coded will become a high-dimensional sparse matrix [19]. Next, data normalization and descriptive statistics were performed on the processed string and numerical data, and the data was segmented into a test and a training set in a 1:4 ratio. Feature engineering includes feature processing and feature selection. Before establishing a prediction model, the first step is to standardize the features, scale the original data proportionally, and map it to the inter cell range. The expression of the Max Min standardized transformation is Eq. (1).

$$x = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

In Eq. (1), after feature processing, the numerical data is uniformly normalized to the range of [0,1] intervals. The technical process of feature selection is Fig. 3.



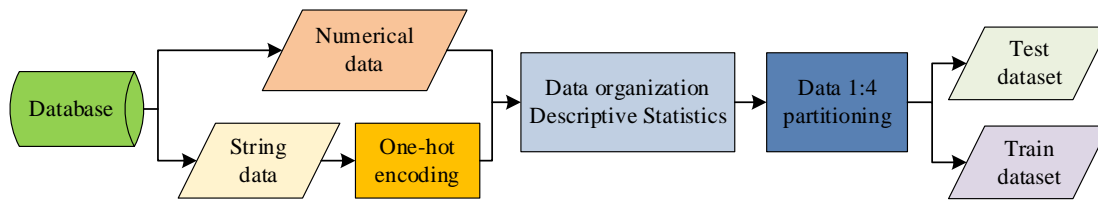


Fig. 2. Data preprocessing flowchart.

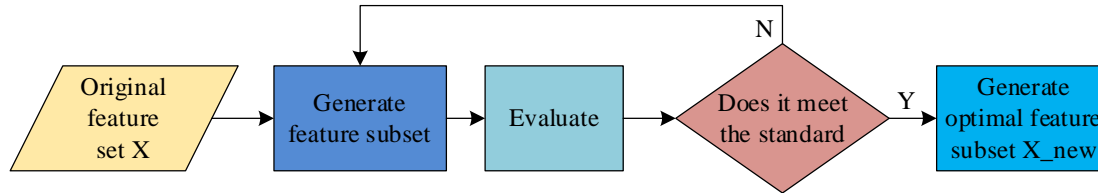


Fig. 3. Feature selection technical process.

In Fig. 3, a part of the feature subset is generated from the original feature set, and the optimal feature subset is generated by evaluating the selected subset that meets the criteria. If the criteria are not met, continue selecting in the feature subset. In the process of calculating advertising click logs, there are extremely few cases where users click on the logs, with the vast majority being cases where users have not clicked. There is a problem of category imbalance in CTR prediction. Category imbalance refers to the significant difference in the number of training samples for sample data in classification tasks. To avoid imbalanced class distribution, data sampling uses an under sampling method that removes the number of multiple class samples. The CTR prediction results only have two possibilities: clicking or not clicking, which belongs to the binary classification problem: user clicks are represented by  $y=1$ , and user clicks are represented by  $y=0$ . Table 1 shows the binary confusion matrix on the obtained test set.

TABLE I. DICHOTOMOUS CONFUSION MATRIX

Predictive Value \ True Value	Regular Class ( $y=1$ )	Negative Class ( $y=0$ )
Regular class ( $\hat{y}=1$ )	TP	FN
Negative class ( $\hat{y}=0$ )	FP	TN

TP represents the amount of correctly predicted positive samples. FP is the prediction error numbers for negative examples. FN represents the quantity of prediction errors for active samples. TN refers to the correct negative example amounts predicted. From the confusion matrix, the total evaluation index of the method is Eq. (2).

$$accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad (2)$$

In Eq. (2), Accuracy, Precision and Recall rate are all used as measurement indicators. Precision is the possibility of true prediction in the positive sample predicted by the model, expressed as follows.

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

The recall rate represents the proportion of all active examples, as expressed in Eq. (4).

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

The ROC indicator is connected with True Positive Rate (TPR) and False Positive Ratio (FPR), where the TPR calculation formula is Eq. (5)

$$TPR = TruePositiveRate = \frac{TP}{TP + FN} \quad (5)$$

The FPR calculation formula is Eq. (6).

$$FPR = FalsePositiveRate = \frac{FP}{FP + TN} \quad (6)$$

TPR and FPR are interdependent, and the larger the TPR and the smaller the FPR, the more superior the classification performance. Logloss index of the logarithmic loss function measures the prediction accuracy of the classifier, and can also be used as the standard of the classification effect, as shown in Eq. (7).

$$\log loss = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M y_{ij} \log(p_{ij}) \quad (7)$$

In Eq. (7),  $N$  is the gross samples, and  $p_{ij}$  represents the predicted CTR of users to advertisements. The logarithmic loss of all samples represents the average logarithmic loss of each sample, and the smaller the value, the better the model performance. The evaluation indicator Area Under Curve (AUC) of the method is expressed as the area enclosed by the coordinate axis under the ROC curve. In terms of CTR prediction, the larger the AUC, the better the training effect. The smaller the Logloss value, the higher the accuracy of the prediction model [20].

### B. DNN-based Prediction of Advertising CTR

On the basis of constructing A-CTR-P, further research is conducted on the use of DNN for CTR prediction. DNN is a multi-layer unsupervised neural network that uses the output features of the previous layer as inputs to the next layer for feature learning. After layer by layer feature mapping, the features of existing spatial samples are mapped to another feature space to learn better feature expression for existing inputs. DNN has multiple nonlinear mapping feature transformations that can fit highly complex functions [21]. The structure is Fig. 4.

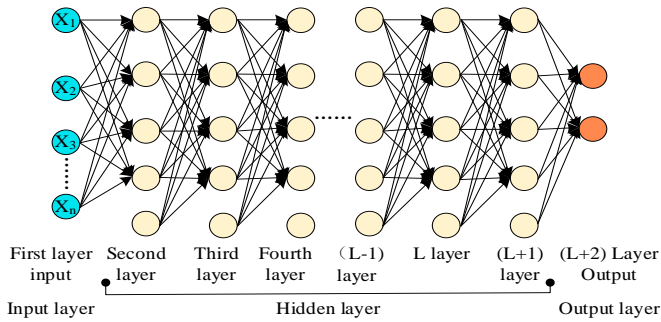


Fig. 4. DNN model structure.

In Fig. 4, DNN includes input, hidden and output layers. In CTR prediction, i.e. the binary classification  $\{0,1\}$  problem, the DNN input layer with an L-layer hidden layer is Eq. (8).

$$X = \{X_1, X_2, X_3, \dots, X_n\} \quad (8)$$

$n$  in Eq. (8) means the input quantities. The hidden layer is Eq. (9).

$$h^{(l)} = f(W^{(l)}h^{(l-1)} + b^{(l)}) (\forall l \in 1, 2, 3, \dots, L-1) \quad (9)$$

In Eq. (9),  $h^{(l)}$  represents the input vector of L+1,  $f(\cdot)$  is the activation function.  $W^{(l)}$  represents the weight matrix of L-1.  $b^{(l)}$  is the offset vector of L. The output layer is Eq. (10).

$$y_{pre} = \arg \max_C P(y = C, X; w, b) \quad (10)$$

In Eq. (10),  $X$  represents the input vector, and  $P(y = C, X; w, b)$  represents the probability that the output is equal to  $C$ .  $y_{pre}$  is the output of the final model and the corresponding category  $C$  when  $P(y = C, X; w, b)$  is at its maximum, resulting in category 0 or 1. The key solution  $\{w, b\}$  of DNN model calculation makes the loss function the smallest, and the DNN weight could be trained through Stochastic Gradient Descent (SGD). SGD is an optimization algorithm used to update DNN parameters on a gradient basis. Each iteration will randomly select a small batch of samples to calculate the gradient of the loss function, and use the gradient to update the parameters. This random characteristic makes the algorithm more robust, avoiding getting stuck in local minima, and also accelerates training speed [22]. Random

gradient descent is the average extraction of a small batch of samples  $B = \{x^{(1)}, \dots, x^{(m)}\}$  from the training set.  $m$  is usually a relatively small number. When the amount of data is large, it is necessary to iterate with abundant samples to gain the optimal solution, as expressed in Eq. (11).

$$g = \frac{1}{m} \nabla_{\theta} \sum_{i=1}^m L(x^{(i)}, y^{(i)}, \theta) \quad (11)$$

Eq. (11) uses samples from a small batch of  $B$ . To obtain the minimum cost function and the optimal parameters, the weight update rules used in the training of DNN are Eq. (12).

$$\varpi^{(r+1)} = \varpi^{(r)} - \nabla E(\varpi^r) \quad (12)$$

In Eq. (12),  $\varpi$  is the weight,  $\nabla$  is the gradient, and  $E$  is the error function. The selection of activation function includes Sigmoid and Rectified Linear Unit (Relu). In the sigmoid function, when any input value is  $x$ , the output value is between the intervals (0, 1). When the input value is 0, the output value is 0.5. The sigmoid function can be represented as Eq. (13).

$$f(x) = \frac{1}{1 + \exp(-x)} \quad (13)$$

The Relu function is Eq. (14).

$$g(x) = \max(0, x) \quad (14)$$

The derivative of  $g(x)$  can be obtained from Eq. (14), which can be gained as Eq. (15).

$$g(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (15)$$

In Eq. (15), if the input value  $> 0$ , the output is equal to it; If it  $< 0$ , the output = 0. DNN can handle high-dimensional sparse category features, but its ability to learn samples with multiple parameters, longer training time, and fewer categories is limited. SDNN is a research on random undersampling of data based on DNN, in order to improve the noise and imbalance of the data. Random undersampling can improve runtime and solve storage problems by reducing the number of samples. The SDNN structure is listed in Fig. 5.

In Fig. 5, the input numeric data and string data are normalized and heat coded respectively to obtain a normalized numeric matrix and a sparse matrix. The two are connected by a matrix to form training data. The training data will be normalized using resampling techniques. Input the resampled data for deep feature learning of DNN, and then output the predicted value probability and corresponding predicted labels. SDNN is a new method improved in DNN. After normalizing and encoding the input data, a random under sampling process is added to the data, thereby improving the impact of data imbalance on DNN and mining complex association relationships in features. The process framework of the SDNN algorithm is Fig. 6.

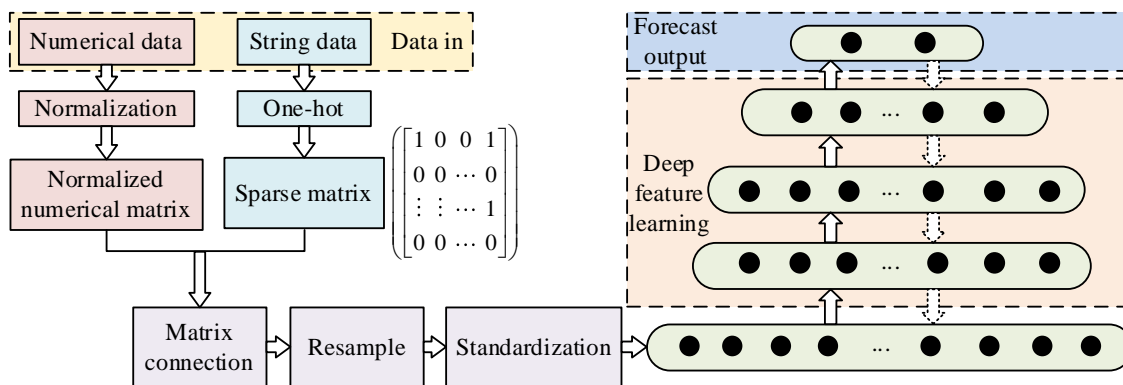


Fig. 5. SDNN structure.

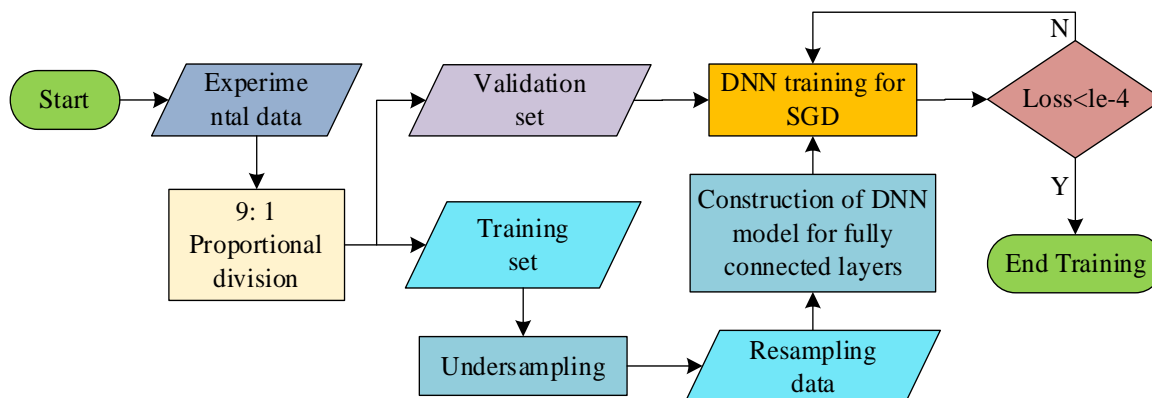


Fig. 6. SDNN algorithm process.

In Fig. 6, the input training data is separated into a training and a validation set in 9:1. Before constructing the DNN, a random under sampling technique is taken to the training set, and then a fully connected DNN is established on the resampled dataset. In training, the loss function on the verification set needs to be calculated. When the number of iterations increases and the loss function no longer changes or changes very little, that is,  $loss < 1e-4$ , the training ends. For parameter issues, the study adopts the dropout technique proposed by Hinton. Dropout can ensure that weight updates no longer rely on the joint action between hidden layer nodes. Overall, the network structure undergoes changes during each DNN training session. The general effect of the final model depends on the synthesis of different model predictions each time.

#### IV. ANALYSIS OF A-CTR-P RESULTS BASED ON DNN

Starting from DNN, to construct a prediction model and first conduct model exploration experiments based on DNN. To comprehensively verify the effectiveness of the DNN, the DNN model and the improved model SDNN were studied and designed. The activation function ReLU and Sigmoid were compared and analyzed, and the key parameter dropout was effectively explored. First, a comparative analysis of the activation function ReLU and Sigmoid is carried out, as listed in Fig. 7.

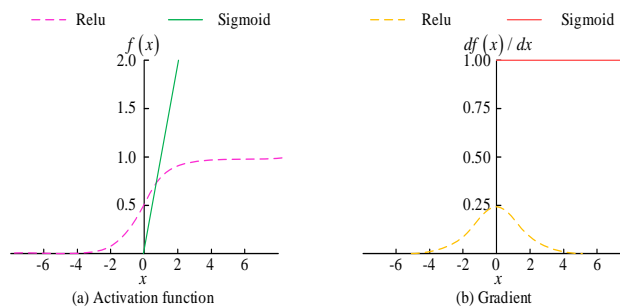


Fig. 7. ReLU and sigmoid activation function and gradient comparison diagram.

In Fig. 7(a), the input value of the sigmoid activation function is less than 0 and approaches 0, while the output value is greater than 0 and approaches 1. In the process of backpropagation, only when the input is around 0 has good activation. In Fig. 7(b), the sigmoid function makes the neural network better at feature recognition, but generally causes the gradient to disappear within five layers. The ReLU activation function is constant at gradients greater than 0, effectively avoiding the matter of gradient disappearance. The gradient of the ReLU function is relatively stable compared to the sigmoid function, indicating that the AUC value of the ReLU is relatively stable. To study the feature learning ability of hidden layers, the effects of the model structures of ReLU DNN/SDNN and Sigmoid DNN/SDNN on AUC were compared and analyzed, as displayed in Fig. 8.

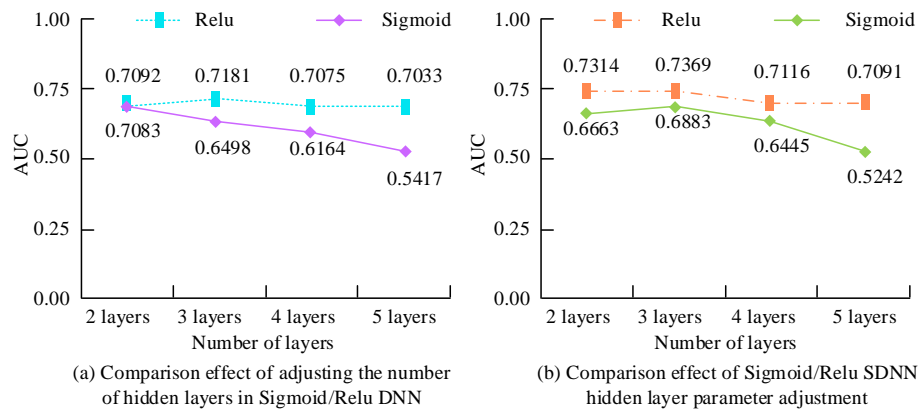


Fig. 8. Comparison rendering Relu/sigmoid activation function and gradient comparison diagram.

In Fig. 8(a), the AUC values of Sigmoid and Relu are similar when the hidden layers are 2. The AUC curve of Sigmoid DNN gradually deducts as the hidden layers increases. When the hidden layer amounts are 5, the AUC value is the lowest, at 0.5417. The AUC values of Relu DNN are all around 0.70, and the curve is relatively stable. The AUC reaches its peak at 0.7181 when the layers are 3. The AUC of Relu DNN has significantly lifted, and Relu represents the relative stability of the prediction effect. In Fig. 8(b), until the layers are 5, the AUC value of Sigmoid SDNN is the lowest, at 0.5242. The AUC value of Relu SDNN is relatively stable, reaching a peak of 0.7369 when the number of hidden layers is 3. Under the same model structure, the AUC value of SDNN is higher than that of DNN, indicating that DNN random under sampling helps to improve the AUC value. The Relu SDNN model was selected for the study, and the parameter settings are Table II.

TABLE II. DNN PARAMETER SETTINGS

Items	Settings
Model structure	2022-1024-1024-800-2
Objective function	Mean_squared_error
Max-iterations for training	200
Activation function	Relu
Regularizer	L2

To ensure the effectiveness of the algorithm, research was conducted under GPU parallel acceleration. The training time of the DNN model is 277.1801s, and the SDNN is 74.1490s, which is approximately four times that of the SDNN. The training time of SDNN has been reduced by about 73.25%, greatly improving the computational efficiency of DNN. DNN trained on the original data and reduced the data samples by 60% after under sampling. The training scale of SDNN has decreased. However, the reduction of training data did not make the prediction performance of SDNN worse; on the contrary, it also improved the computational efficiency of the algorithm. Considering that the number of network layers and the number of iterations in the training phase of SDNN also have an impact on the prediction results of the model, in order to obtain more reasonable parameter values, the experiment was trained on a training set with a data size of 400000 samples. In the experiment, different network layers and

iterations were selected to obtain the AUC values of the model, as shown in Table III.

TABLE III. AUC VALUES OF THE MODEL UNDER DIFFERENT NETWORK LAYERS AND ITERATIONS

Number of layers	Iterations					
	50	70	90	110	130	150
2	0.8103	0.8213	0.8237	0.8251	0.8248	0.8242
3	0.7821	0.8103	0.8194	0.8215	0.8202	0.8214
4	0.8215	0.8327	0.8392	0.8482	0.8501	0.8503
5	0.7979	0.8267	0.8372	0.8380	0.8376	0.8371
6	0.8064	0.8132	0.8158	0.8263	0.8257	0.8278

In Table III, regardless of the number of iterations, the AUC of the model is the highest when the number of network layers is 4. Regardless of the number of network layers, when the number of iterations is less than 110, the AUC value of the model remains increased. When the number of iterations is 110 or above, the AUC value of the model is relatively stable and does not change much. To determine the values of the key parameter dropout in the DNN model, the study was conducted from 0.2 to 0.9 in steps of 0.1. The impact of adjusting the dropout parameter on the prediction performance of SDNN, as well as the comprehensive comparison of the four models with a dropout of 0.5 is exhibited in Fig. 9.

In Fig. 9(a), as the dropout increases, the AUC curve gradually rises and gradually decreases after reaching its peak. When the dropout is greater than 0.8, the AUC curve sharply decreases. When the dropout is 0.5, AUC reaches its peak at 0.7394. When the dropout is less than 0.8, the AUC curve is relatively stable. Therefore, the parameter dropout value is set to 0.5. In Fig. 9(b), the AUC value of DNN is 0.7092, the AUC value of DNN model trained with dropout is 0.7300, and the AUC value of SDNN is 0.7207. The SDNN model trained with dropout has the highest AUC, which is 0.7394. The AUC value of SDNN has been improved, indicating that resampling can achieve the goal of eliminating data imbalance. Balanced data categories have a certain effect on improving prediction performance. As a result of the large size of the data itself, the SDNN was selected for comparative analysis between the training and the test set. When the dropout parameter is 0.5, the training iteration process of SDNN is Fig. 10.

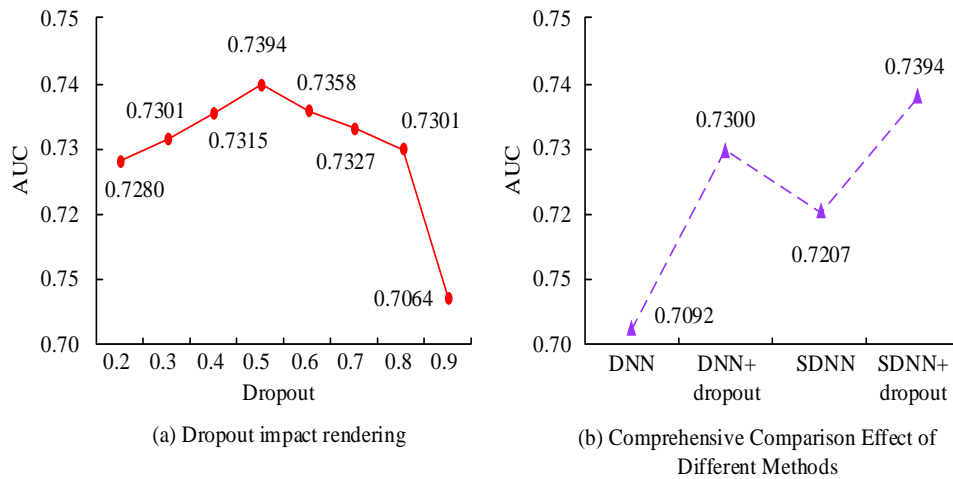


Fig. 9. Comprehensive comparison chart.

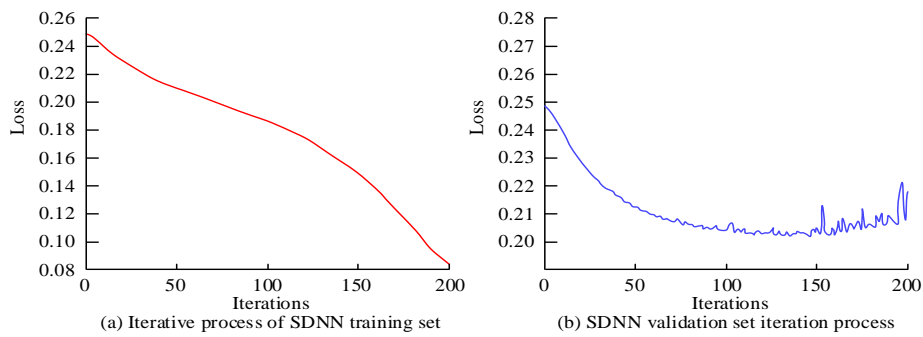


Fig. 10. The training iteration process of SDNN.

In Fig. 10(a), when the number of iterations of the SDNN model training set trained with dropout is 200, the loss function curve still shows a downward trend, and the loss function is 0.085. In Fig. 10(b), when the number of iterations of the verification set is the 110th, the loss function converges, and the value of the loss function is 0.201. The number of iterations decreases rapidly before 110, and after 110, the curve oscillates and no longer decreases. SDNN is effectively trained. To evaluate the predictive precision, a comparative analysis was performed on four models: DNN, SDNN, DNN trained with dropout, and SDNN trained with dropout. The AUC and Logloss values of the four methods are listed in Fig. 11.

In Fig. 11(a), the SDNN trained with dropout has the highest average AUC, reaching convergence at 110 iterations with an AUC of 0.7375. The DNN trained with dropout

converges at 130 iterations, with an AUC of 0.7300. SDNN converges at 120 iterations, with an AUC of 0.7260. The average AUC of DNN is the lowest, and it gradually converges when the number of iterations is 140, at which point the AUC is 0.7255. In Fig. 11(b), the Logloss curve of the SDNN trained with dropout converges at 110 iterations, with a Logloss value of 0.201. When the number of iterations is 110, the Logloss value of DNN is 0.208, the Logloss value of SDNN is 0.207, and the Logloss value of DNN trained with dropout is 0.203. At this point, the Logloss value of SDNN decreased by 0.4% compared to DNN. The application effects of the four models in the test training set are represented by ROC curves. At the same time, to compare the relationship between the ROC curves of various models more clearly, the area of curve FPR0.2 to 0.3 is enlarged. The comparison of ROC curves and ROC partial curves of the four models are demonstrated in Fig. 12.

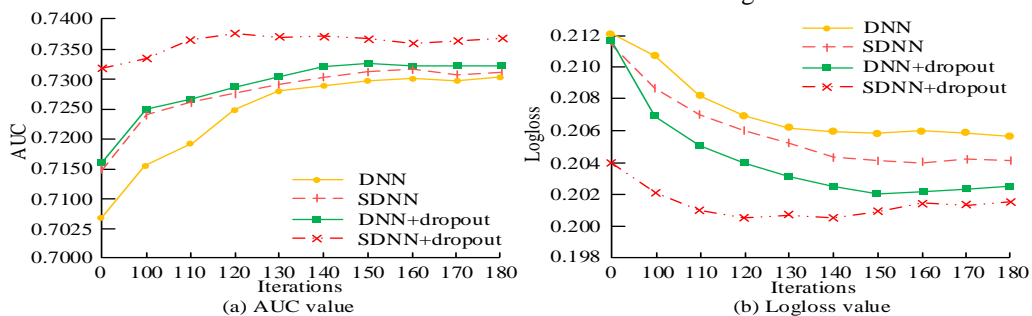


Fig. 11. Comparison of convergence rates under different models.

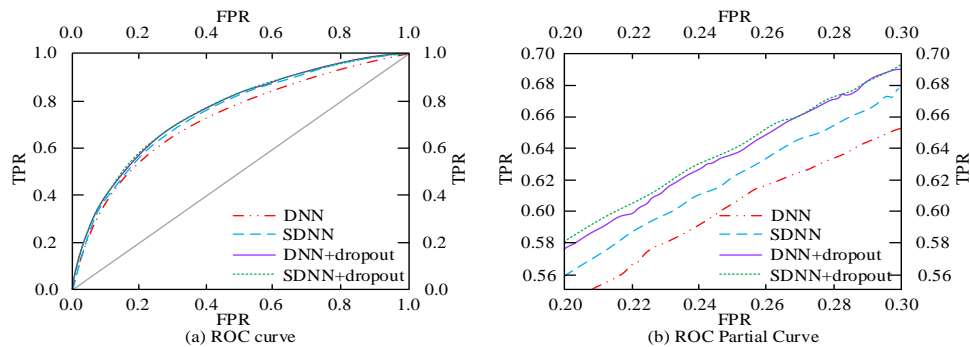


Fig. 12. Comparison between ROC curve and ROC local curve.

In Fig. 12(a), the area under the ROC of DNN is the smallest, at 0.7092, indicating that its AUC is the smallest among the four models. The area under the ROC curve of DNN and SDNN trained with dropout is similar to that of SDNN trained with dropout. In Fig. 12(b), the AUC value of DNN trained with dropout is 0.7300, SDNN trained with dropout is 0.7394, and SDNN is 0.7207. Compared to DNN, SDNN increased AUC by 1.6%. Although the improvement is not significant solely in terms of numerical values, the CTR prediction problem of internet advertising companies involves massive amounts of data every day. Due to the difficulty in estimating costs, any slight increase in AUC value will result in significant advertising effectiveness.

## V. DISCUSSION

The complexity of the model is related to its performance. The higher the complexity of the model, the more accurate the error of the training set at the end of the month, and the resulting model accuracy. However, higher complexity may lead to an increase in the computational workload and training time of the model, leading to an increase in research costs. Therefore, it is necessary to study and select the appropriate number of network structure layers and iterations. The study selects a network layer of 4, iteration number of 110, and parameter dropout value of 0.5. Under these conditions, the model studied can improve the predictive accuracy of advertising placement in a targeted manner, which plays a certain role in making marketing decisions for enterprises.

## VI. CONCLUSION

To improve the prediction accuracy of advertising placement, an improved DNN based A-CTR-P: SDNN was studied and designed, while the activation function and parameter dropout were determined. The prediction accuracy of DNN, DNN trained with dropout, SDNN, and SDNN trained with dropout were compared and analyzed. The results indicate that when the hidden layers have 5, the AUC of Sigmoid DNN is the lowest, at 0.5417. When there are 3 hidden layers, the AUC of Relu DNN reaches its peak at 0.7181, and the AUC is relatively stable at around 0.70. The AUC value of Relu DNN has greatly lifted, and Relu represents the relative stability of the prediction effect. Under the same model structure, the AUC value of SDNN is higher than that of DNN, and the improvement effect of SDNN on DNN is conducive to the improvement of AUC. When the dropout is around 0.5 and the SDNN prediction model training

set has 200 iterations, the loss function curve still maintains a downward trend, at which point the loss function is 0.085. The loss function curve does not decrease until the 110th time of the validation set, at which point the function value is 0.201. When the number of iterations is 110, the Logloss value of DNN is 0.208, and the Logloss value of SDNN is 0.207. At this time, the Logloss value of SDNN decreases by 0.4% compared to DNN. Comparing the ROC curves of DNN and SDNN, the AUC of DNN is 0.7092 and that of SDNN is 0.7207. The prediction accuracy of SDNN is higher than that of DNN, with an increase of 1.6% in AUC. This research result can accurately target users and save the operating costs of the advertising placement system, providing a new direction for advertising companies' marketing and effectively ensuring corporate profits. One of the future research directions is to integrate user social information into advertising CTR prediction models, enrich user information, and improve user expression information. However, due to equipment limitations and limited training data, the advantages of deep learning cannot be fully utilized. So in the future, the amount of training data will be increased so that the model can perform better.

## REFERENCES

- [1] Wang Q, Liu F, Huang P, Xing, S., & Zhao, X. A Hierarchical Attention Model for CTR Prediction Based on User Interest. *IEEE Systems Journal*, 2020, 14(3):4015-4024.
- [2] Erico C, Salim R, Suwanto S, & Isa, S. M. Mobile Advertisement Click through Rate Prediction. *MATTER: International Journal of Science and Technology*, 2020, 29(4s):1534-1540.
- [3] Fan B, Fan W, Smith C, & Garner, H. S. Adverse drug event detection and extraction from open data: A deep learning approach. *Information Processing & Management*, 2020, 57(1):102131.1-102131.14.
- [4] Lyu B, Hu Y, Zhang W, Du, Y., Luo, B., & Sun, X. Fusion Method Combining Ground-Level Observations with Chemical Transport Model Predictions Using an Ensemble Deep Learning Framework: Application in China to Estimate Spatiotemporally-Resolved PM<sub>(2.5)</sub> Exposure Fields in 2014-2017. *Environmental Science & Technology*, 2019, 53(13):7306-7315.
- [5] Feng C, Zhang H, Wang S, Li, Y., Wang, H., & Yan, F. Structural Damage Detection using Deep Convolutional Neural Network and Transfer Learning. *KSCE journal of civil engineering*, 2019, 23(10):4493-4502.
- [6] Cai W, Wang Y, Ma J, & Jin, Q. CAN: Effective cross features by global attention mechanism and neural network for ad click prediction. *Tsinghua Science and Technology*, 2021, 27(1): 186-195.
- [7] Xue N, Liu B, Guo H, Tang, R., Zhou, F., Zafeiriou, S., ... & Li, Z. Autohash: Learning higher-order feature interactions for deep ctr prediction. *IEEE Transactions on Knowledge and Data Engineering*, 2020, 34(6): 2653-2666.

- [8] Liu Y, Pang L, Lu X. Click-through Rate Prediction Based on Mobile Computing and Big Data Analysis. *Ingénierie des Systèmes D Information*, 2019, 24(3):313-319.
- [9] Zhou L. Product advertising recommendation in e-commerce based on deep learning and distributed expression. *Electronic Commerce Research*, 2020, 20(2): 321-342.
- [10] Ghorbel A, Souissi B. Upper confidence bound integrated genetic algorithm-optimized long short-term memory network for click-through rate prediction. *Applied Stochastic Models in Business and Industry*, 2022, 38(3):475-496.
- [11] Liu Y, Yang T, Qi T. An Attention-Based User Preference Matching Network for Recommender System. *IEEE Access*, 2020, 8(99):41100-41107.
- [12] Hung J L, Rice K, Kepka J, & Yang, J. Improving predictive power through deep learning analysis of K-12 online student behaviors and discussion board content. *Information Discovery and Delivery*, 2020, 48(4):199-212.
- [13] Ma J, Shang P, Zou X, Ma, N., Ding, Y., & Su, Y. Remaining Useful Life Transfer Prediction and Cycle Life Test Optimization for Different Formula Li-ion Power Batteries Using a Robust Deep Learning Method. *IFAC-PapersOnLine*, 2020, 53( 3):54-59.
- [14] Ozmen E P, Ozcan T. A novel deep learning model based on convolutional neural networks for employee churn prediction. *Journal of Forecasting*, 2022, 41(3):539-550.
- [15] Deng H, To A C. Topology optimization based on deep representation learning (DRL) for compliance and stress-constrained design. *Computational Mechanics*, 2020, 66(2):449-469.
- [16] Hu B, Li J. An Edge Computing Framework for Powertrain Control System Optimization of Intelligent and Connected Vehicles based on Curiosity-driven Deep Reinforcement Learning. *IEEE Transactions on Industrial Electronics*, 2021, 68(8):7652-7661.
- [17] He H, Jiang H. Deep Learning Based Energy Efficiency Optimization for Distributed Cooperative Spectrum Sensing. *IEEE Wireless Communications*, 2019, 26(3):32-39.
- [18] Polvara R, Sharma S, Wan J, Manning, A., & Sutton, R. Autonomous Vehicular Landings on the Deck of an Unmanned Surface Vehicle using Deep Reinforcement Learning. *Robotica*, 2019, 37(11):1867-1882.
- [19] Jafarzadehpour F, Molahosseini A S, Zarandi A, & Sousa, L. Efficient Modular Adder Designs Based on Thermometer and One-Hot Coding. *IEEE transactions on very large scale integration (VLSI) systems*, 2019, 27(9):2142-2155.
- [20] Heo S K, Nam K J, Loy-Benitez J, Li, Q., Lee, S. C., & Yoo, C. K. A deep reinforcement learning-based autonomous ventilation control system for smart indoor air quality management in a subway station. *Energy and Buildings*, 2019, 202(Nov.):109440.1-109440.16.
- [21] Wang X, Cheng M, Eaton J, et al. 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.

# Design and Development of an Intelligent Rendering System for New Year's Paintings Color Based on B/S Architecture

Zaozao Guo

Faculty of Art, Sustainable and Creative Industry, Sultan Idris Education University, 35900, Perak Darul Ridzuan, Malaysia  
Department of Computer Science, Nankai University Binhai College, Tianjin 300270, China

**Abstract**—With the arrival of the synthetic talent era, laptop technological know-how for the safety and inheritance of intangible cultural heritage has added a new way of thinking, and the range of intangible cultural heritage additionally offers greater chances for laptop technology, the utility of laptop talent science to New Year's Eve artwork of the applicable lookup there are many gaps. Training of Cyclic Generative Adversarial Network (CycleGAN) realize the task of extracting plots of different site types from planar maps and the Rendering generation from planar color block map to color texture map. This paper first introduces the B/S community architecture, Python programming technological know-how and Django framework. Then the unique approach of using pc Genius to the project of rendering Chinese New Year artwork is clarified via modeling, studying algorithms, and community architecture. Finally, a hierarchical fusion generative adversarial neural community structure is designed primarily based on generative adversarial neural networks. The structural and textural features of the image are fused by texture GAN and then rendered to generate the New Year paintings. The test results show that this kind of algorithm draws clear texture, realistic images and full color of the New Year's pictures, and the IS index reaches 3.16 in the quantitative analysis, which is higher than other comparison algorithms.

**Keywords**—B/S architecture; intelligent rendering; adversarial neural network; Chinese New Year painting

## I. INTRODUCTION

There are two essential standard approaches to instructing painting: one is the way of offline teaching, and the different is the way of online video teaching. Offline introduction is an expert instructor and college students face-to-face teaching; this offline coaching is the most common way of educating artwork education [1-3]. The online video instructing technique is that the instructor documents the video in advance, and then the college students pay to get the instructing video. The two present regular portray teaching strategies have the following disadvantages.

1) *Offline* teaching methods of one-on-one teaching is the teacher to meet the requirements of the students at the designated location of the class; this class is expensive; if it is a one-to-many teaching mode, although the cost can be relatively reduced, the teaching effect cannot be guaranteed;

2) *The* relevant video teaching method is the teacher to meet the requirements of the students to the designated location of the class.

3) *Fewer* relevant professional teachers, but the market demand, resulting in several students still in the school stage for unregulated training courses, resulting in uneven teaching levels.

4) *Teachers* and students offline teaching methods by space, time limitations, teachers and students or student's parents must be based on the designated time to the designated place of class teaching.

5) *The* traditional video teaching method is usually that the teacher records the video in advance, and then the students watch the video to learn. This teaching method lacks interaction between the teacher and the students and is passive.

Painting is one of the oldest types of art, and its content material has advanced all through human records and exclusive cultures. In modern times, with the popularization of electronic information technology, all kinds of electronic drawing board software and hardware have gradually become the main tools for people to create paintings, such as Adobe's Photoshop, Apple's iPad, Apple Pencil, etc., and the digital form of photographs produced by them have been widely used in various aspects of production and life, such as news dissemination, prototyping, film and television creation, etc. In general, portraying wholly demonstrates human nature, and it is the essential painting shape, focusing on realism. In general, representation wholly shows human knowledge and creativity and is a critical potential in visible conversation for humans regardless of its shape and tools. In the modern-day community era, the digital shape of portray can be very handy to disseminate and use, is one of the major types of contemporary media, and consequently, has a very realistic fee and realistic significance [4, 5].

The main manifestation of program communication, the canonical representation of the floor plan with the advancement of AI technology, the design drawing and high-quality rendering based on the design is crucial for the design presentation [4]. Therefore, machine learning (Machine Learning) of the setup data makes it possible for Artificial Intelligence Aided Design (Artificial Intelligence Aided planner often spends a lot of time to collect plan cases that can



be drawn from, Design, AIAD), which, Deep Learning and rendering of the floor plan using a variety of software. At present, computer vision (Deep Learning) has been in the design of image analysis and generation of consciousness (Computer Vision) in the field of image recognition; generation technology shows great potential for application. In recent years, in architecture and graphic arts has been more mature, so that AI recognition case plane and rendering cartography design field, professional image data sets and deep learning applications do not become possible. And the realization of this automated analysis and mapping of the front break emerges, designers rely on interdisciplinary cooperation to develop AIAD field mentioning is a high-quality landscape garden floor plan training set. Unlike common face or object image databases, Landscape Architecture Flat View to improve the efficiency of design analysis and mapping. And landscape architecture discipline to carry out this type of research limitations mainly from the interdisciplinary cooperation and the number of face database has outstanding professionalism: 1) the variety of land types: the lack of data sets. 2) drawing expression both normative and artistic; 3) access to the threshold of the image is to present the design scheme is an important carrier of the Landscape Garden high, the need for professionals to carry out careful screening and data annotation Lin design floor plan is the designer case to draw upon, Concept presentation and Data Labeling.

Artificial intelligence is an important engine for future economic development, and machine learning technology is a major research field of artificial intelligence; the current machine learning technology, according to the different learning modes, is roughly divided into three categories: supervised learning, unsupervised learning and reinforcement learning. Supervised gaining knowledge is a studying technique that depends on statistics labeling, primarily the use of statistics labeling statistics to supervise the coaching model and then using the dummy to predict the new data, in general, used in classification, regression and different tasks [6]. In general, supervised and unsupervised learning focus on the perception and understanding of data; in contrast,

As shown in Fig. 1, the overall architecture of Python can be divided into three parts: Python module libraries, commonly used standard libraries and runtime environments. For the commonly used syntax standard libraries of Python, (see Table I).

TABLE I. COMMON SYNTAX STANDARD LIBRARY OF PYTHON

Grammar Library	Related Explanations
atexit	Functions that are allowed to be called when the program exits
argparse	A function that parses a command run
bisect	Providing a binary lookup algorithm for sorted lists
calendar	Date-related functions
codecs	Functions related to coding and decoding data
collections	Useful data structures
copy	Functions for copying data

reinforcement learning methods not only realize the perception and understanding of data but also focus on generating intelligent decisions and actions according to the task goals, which can realize more complex functions and become a hot research topic in recent years.

## II. RELEVANT THEORIES AND TECHNOLOGIES

Technology is the basis for realizing various functions. In building a platform for processing and analyzing painting images, some important theoretical techniques provide important support for the platform's design. This paper adopts the Python+Django technology framework for design and uses PyCharm for design and debugging. The B/S architecture is used for system design to facilitate subsequent upgrading and maintenance [7, 8]. Before introducing the platform, the theoretical technologies involved are first introduced and summarized.

### A. Python Technology

Python is a high-level and broadly used programming language. It has the benefits of excessive efficiency, low overhead, open source, portability, sturdy interpretability, etc., and has been preferred to utilize builders quickly. Python's fusion of several algorithmic programs determines that it can play a necessary position in unique fields. The general architecture of Python's distribution of functions is shown in Fig. 1 as follows:

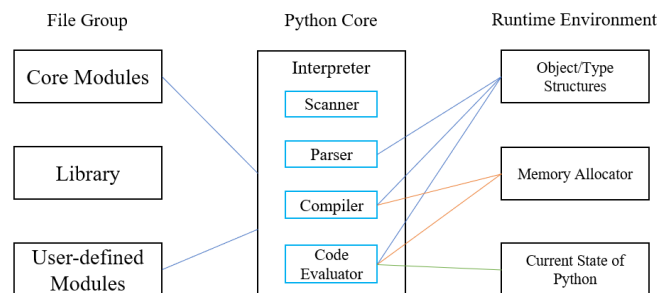


Fig. 1. Python overall architecture diagram.

### B. Cornerstone Image Support Technology

Cornerstone.js is a medical image viewing tool written in Js script to support the display and interaction of medical images. As its name suggests, it has a “cornerstone” role in the image display field, and many image reading systems for artistic paintings are based on Cornerstone.js. Rich scripts provide powerful technical support for browsing image data. Basic operations such as zooming and panning of image data can be realized.

Currently, most home portray exhibition structures mixed with AI are also developed primarily based on Cornerstone.js. The Internet web page multi-threaded decoding used in this script quickens the show of snapshots and helps the Internet to practice compression strategies such as JPEG to transmit images. Modularity (component design) allows it to be embedded in extraordinary front-end frameworks for convenient invocation through developers. As proven in the following code, it can recognize the show of DICOM portray images [9, 10].

```
const
element=document.getElementById('demo-element');
const imageUrl='http://example.url.com/image.dcm?';
cornerstone.enable(element);
cornerstone.loadAndCacheImage(imageId).then(function(i
```

```
mage){
cornerstone.displayImage(element,image);
cornerstoneTools.mouseInput.enable(element);
```

The flow when Cornerstone performs image display is shown in Fig. 2.

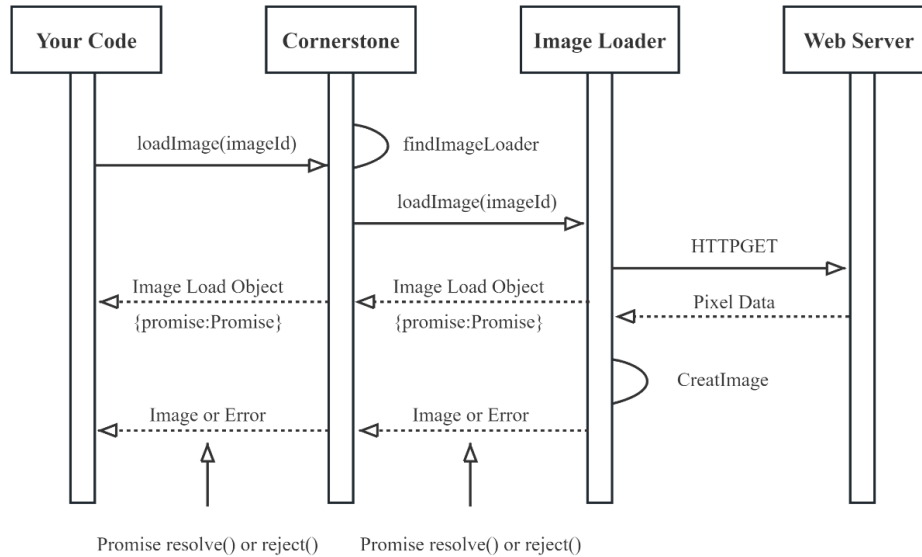


Fig. 2. Cornerstone flowchart.

C. Web Framework Technology

Web framework technological know-how is a disbursed utility software architecture, which is ordinarily divided into two foremost parts: client-side and server-side. The customer aspect frequently includes Html language, script program, CSS style, plug-in technological know-how and so on [11].

Each element on the “active” page has a variety of labels. web server development technology is also from static to dynamic gradual development, gradually being perfected. Server-side technologies include servers, CGI, ASP, Servlet and JSP technologies [12, 13]. Fig. 3 shows the Web application processing flowchart.

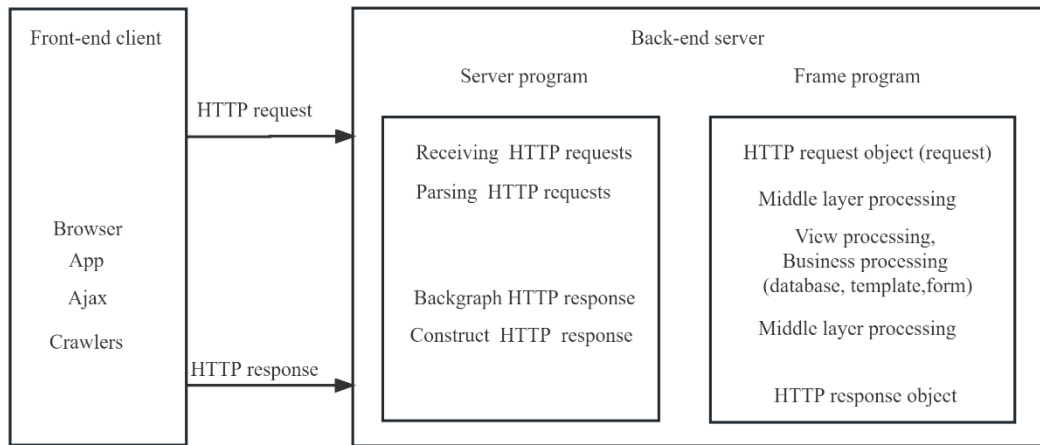


Fig. 3. Web application processing flow chart.

Web technology is more in line with the current technical requirements and is an important direction for the future development of network architecture. Therefore, this paper adopts the B/S architecture model that satisfies Web technology for the research and development of image processing systems [13].

D. MVT Framework

Generally, in improving the project, it is fundamental to first classify the functions, a massive piece of the task damaged down into many small tasks, and as a way as viable to comprehend the low coupling of the range of modules to beautify the scalability and portability.

MVC's full spelling for Model-View-Controller, after continuous development and integration, the idea of MVC has been applied to Web development. It plays an important role, known as Web MVC framework. In MVC, M is model (business logic layer), V is view (interface layer), C is controller (controller), used to schedule View layer and Model layer.

Python language Django framework uses the MVT architecture pattern. In Django for web development, this framework also follows the MVC idea. However, in Django, this architecture is called MVT, but the essence of the MVC pattern is the same.

The operational logic between the three MVTs is shown in the following Fig. 4.

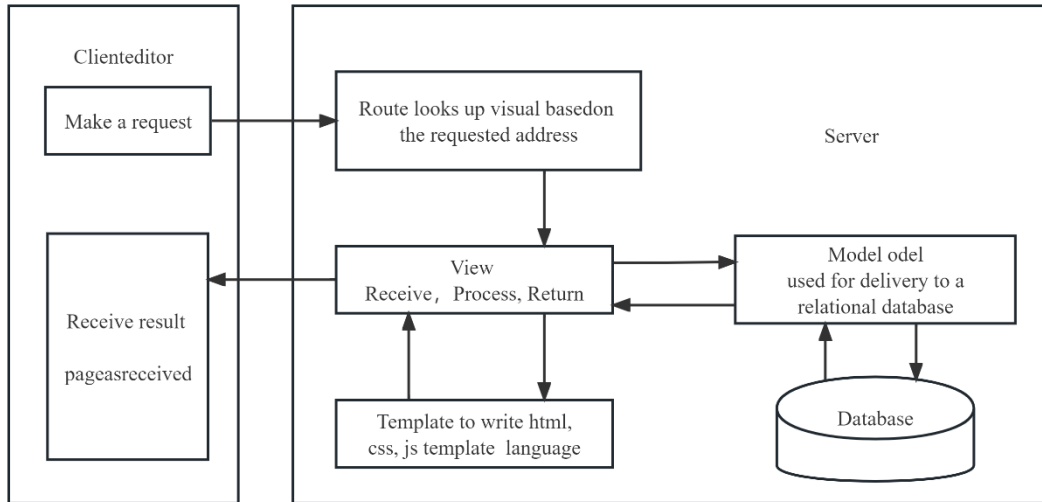


Fig. 4. MVT operation logic diagram.

Various supporting technologies are indispensable in the design and development, and the intermingling of different technologies lays a cornerstone foundation for the design and development of the platform [14]. Coordinating the different technologies to achieve the desired effect is an important and complex process. The role of different technologies are skillfully combined, in line with the needs of the current form of social development, to provide the necessary technical support for the exchange and integration of different fields.

#### E. Architecture Mode

The b/s structure is a famous community shape mode after the upward job of the Web. b/s shape comprises the show, characteristic, and statistics layers. This mannequin unifies the patron and centralizes the core implementation of the device on the server side, simplifying improvement and maintenance [15,16]. The statistics layer strategies and calculates the request based on the user's conduct.

The network structure of the b/s model is shown in the following Fig. 5.

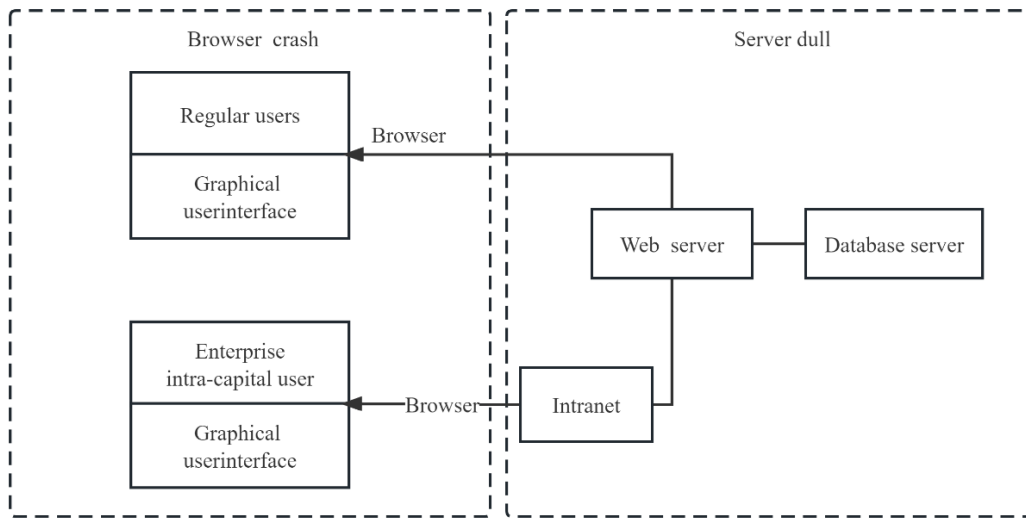


Fig. 5. Schematic diagram of the b/s architecture network structure.

The final processed page information or data is returned to the browser, and the browser displays the results to the browser by rendering the information returned by the server. The page information or data is returned to the browser, and the browser, by rendering the information returned by the server to display the results on the page [17].

Compared with the c/s model, the b/s model has more prominent advantages and is, therefore, more popular.

1) *Easy maintenance and upgrading*: In the common c/s model, upgrading the gadget requires upgrading each the customer and the server. After upgrading the software, every consumer desires to improve his/her purchaser to use it. If widely widespread improvements are required, the work of gadget directors in retaining the gadget will be very time-consuming and poorly maintained. But b/s structure solely wants to focus on upgrading the machine saved on the server can be, all the purchasers, that is, the browser, do no longer want to do any maintenance.

2) *Superior performance*: With the development of the Web, b/s architecture is widely used, but also promotes the development of Ajax technology, which makes part of the data processing in the client can be carried out, greatly reducing the burden on the Web server, and can realize the page local real-time refreshing, improve the interactive performance of the page.

3) *Low development costs, more options*: Software development based on b/s architecture generally only needs to be installed on a Linux server, so there are many choices of servers and high security. The Linux operating system and the database's connection are free, so this not only greatly reduce the cost of software development but also many choices.

4) *High reusability*: The c/s architecture of the program needs to consider the whole software design from the whole; when there is a problem, or the system needs to be upgraded

and changed, it may need to be redesigned to make a completely new system. Software with b/s architecture can be largely divided into different components by functional modules to achieve high reusability.

When the software is changed from c/s architecture to b/s architecture, the system software no longer needs the developers to specialize in the development of the client software. Still, it only needs to focus on the update of the program to free up the labor force [18]. At the same time, the use of the browser as a client, the development of a more friendly interface, while the newly developed system does not require users to learn from scratch, greatly reduces the user's learning costs.

### III. REINFORCEMENT LEARNING BASED STROKE PAINTING SIMULATION

#### A. Model-based DDPG Algorithm

Fig. 6 shows the general framework of the algorithm, which is generally a Model-based Deep Deterministic Policy Gradient (Model-based DDPG). At the core of the framework is a drawing intelligence, whose goal is to decompose a given target image into a number of strokes, and let these strokes reconstruct the target image on the drawing board by means of a renderer, which is "model-based" in the sense that it utilizes the explicit model of a discriminator and a neural renderer [19, 20]. In order to simulate the human drawing process, this paper will use a sequential Markov decision process to model it: in the inference phase, the intelligent body decides the control parameters (i.e., actions) for the next stroke in each step based on the observed target image and the current state of the drawing board, and the renderer receives the control parameters and draws the strokes on the drawing board; in the training phase, the training samples are randomly sampled from the empirical recall cache, and then the discriminators and neural renderers are used to reconstruct the target image with the strokes [21].

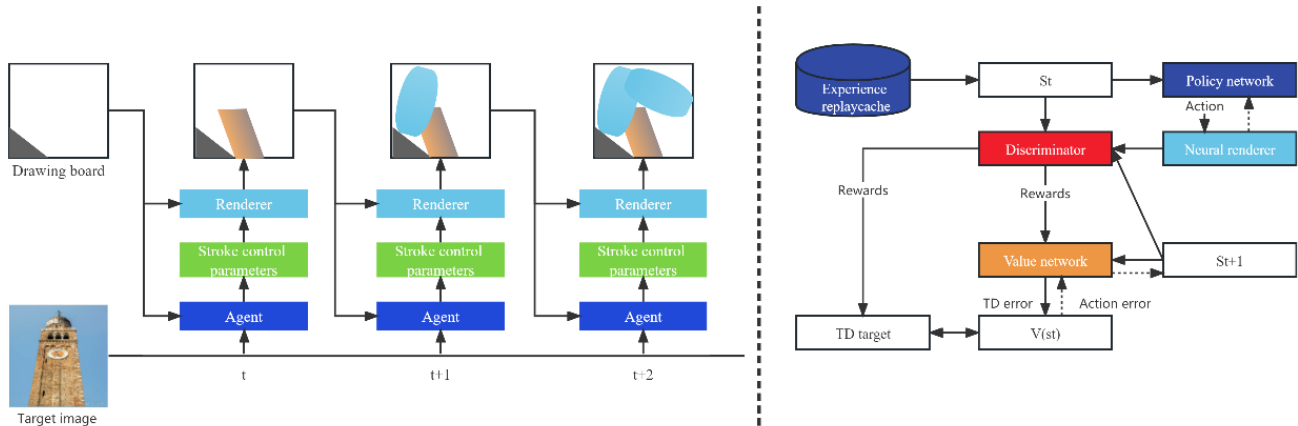


Fig. 6. The general framework of the algorithm (left, inference phase; right, training phase).

#### B. Neural Renderer

1) *Task definition*: In the reinforcement learning drawing simulation framework, the task of the neural renderer is to

render the input stroke control parameters (i.e., the action produced by the drawing intelligences) into a rasterized image of the stroke, which is then added to the current drawing

board. Specifically, neural renderers are used for three reasons.

a) Due to the differentiable nature of the neural renderer, the errors of the strokes can be back-propagated through the renderer, which is crucial for the model-based DDPG algorithm used in this paper;

b) The neural renderer can be trained by a supervised learning algorithm to mimic a real renderer rendering strokes. In this way, existing real renderers can be used to train the neural renderer, avoiding repetitive manual design [22, 23].

2) *Learning algorithm:* In this paper, we use the supervised learning method to train the neural renderer. For each stroke, suppose the corresponding stroke renderer is  $R_{GT}$ , the learning goal of the neural renderer R is to hope that for any drawing board state C and stroke control parameter a,

its rendering result is as similar as possible to  $R_{GT}$ , and ideally, the two are equal, i.e. (as shown in Fig. 7).

$$R(C, a) = R_{GT}(C, a) \quad (1)$$

For the rendering process designed in this paper, the output  $S_t$   $a_t$  of the neural renderer subject network is independent of the drawing board state, taking into account that different strokes have different degrees of prominence in different board states (e.g., black strokes are almost invisible in the black drawing board, while very obvious in the white drawing board). Therefore, during the training process, this paper defines two standard states for the board,  $C_B$ ,  $C_W$ , which denote the empty boards of pure black and pure white, respectively, and then measures the gap in rendering results using the  $L_2$  distance, and ultimately constructs the objective function of the learning as,

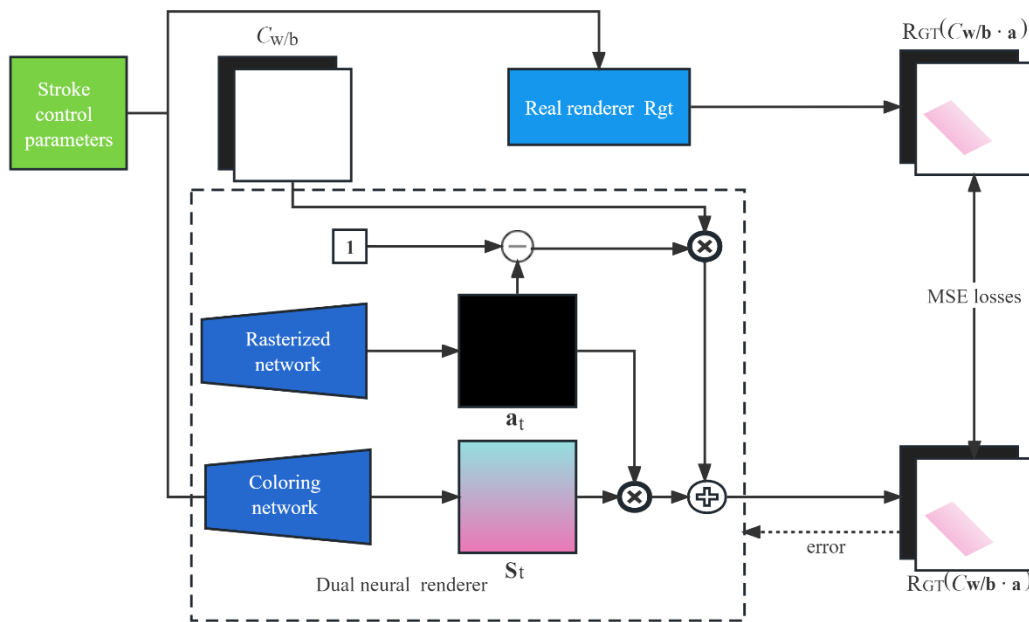


Fig. 7. Neural renderer training approach.

$$J(\mathbf{W}^R) = \mathbb{E}_{a \sim U(A)} \left[ |R(C_B, a) - R_{GT}(C_B, a)|^2 + |R(C_W, a) - R_{GT}(C_W, a)|^2 \right] \quad (2)$$

Where  $\mathbf{W}^R$  denotes the parameters of the neural renderer,  $U(A)$  denotes the uniform distribution defined on the stroke control parameter space A. The learning process of the neural renderer is to minimize the objective function.

$$\min_{\mathbf{w}} J(\mathbf{W}^R) \quad (3)$$

In practice, this paper uses a stochastic gradient descent algorithm to solve this optimization problem and complete the training of the neural renderer. As in Fig. 7, the sampled stroke control parameters will be sent to the real renderer  $R_{GT}$ ,

which renders the stroke on the blank drawing board  $C_{W/B}$  with the result  $R_{GT}(C_W, a)$ , and the neural renderer synthesizes the  $S_t$ ,  $a_t$  internally according to the same control parameters, and then obtains the rendering result  $R_{GT}(C_W, a)$  according to Eq. 3, and finally calculates the  $L_2$  of the two results distance (i.e., MSE loss) to update the network parameters of the neural renderer.

3) *Network structure:* The neural renderer in this paper uses a two-way architecture divided into a shading network and a rasterization network. The shading network is used to output the stroke image  $S_t$ , whose input is the complete stroke control parameters, and the network consists of six layers of inverse convolutional layers. The rasterization network is used to output the transparency image  $a_t$ , whose

input is the part of the stroke control parameter that does not contain color, and the network consists of four fully-connected layers and three Shuffle convolutional layers.

### C. Stroke Translator

1) *Task definition:* The undertaking of the stroke translator is to translate the managed parameters of one variety of strokes into these of any other by means of the capacity of a neural network so that the rendering consequences of the two sorts of strokes are as visually comparable as possible [24]. After the intelligent body has been trained to master one type of stroke, the control parameters of the stroke generated by the intelligent body during the painting process can be collected and then converted into the control parameters of another type of stroke using a stroke translator, and finally rendered by a renderer to obtain the painting results of another type of stroke. In this way, the painting results of multiple styles of strokes can be obtained with less time and computational resources [25, 26]. Let the control parameter of the source stroke A be  $a_A \in A_A$  and the control parameter of the target stroke B be  $a_B \in A_B$ ; the translator  $T_{AB}$  translates the source stroke into the target stroke, i.e.

$$a_B = T_{AB}(a_A; \mathbf{W}^T) \quad (4)$$

Where  $\mathbf{W}^T$  denotes the parameters of the translator.

2) *Learning algorithm:* In this paper, we use a supervised learning method to train a stroke translator. Assuming that the renderer of the source stroke A is  $R_{GT}^A$  and the renderer of the target stroke B is  $R^B$ , where  $R^B$  is a neural renderer, the learning goal of the translator is to hope that for any source stroke  $a_A \in A_A$  and any drawing-board state C, the renderer is able to find a corresponding target stroke  $T_{AB}(a_A; \mathbf{W}^T) \in A_B$  so that the rendering results of the two strokes  $R_{GT}^A(C, a_A)$  and  $R^B(C, T_{AB}(a_A))$  are as similar as possible visually. For this, this paper defines two standard drawing board states,  $C_B$ , and  $C_W$ , which denote pure black and pure white empty drawing boards, respectively, and then measures the gap between the rendering results using the  $L_2$  distance and finally constructs the learned objective function as,

$$J(\mathbf{W}^T) = E_{a_A \sim U(A_A)} \left[ \begin{aligned} & \left| R^B(C_B, T_{AB}(a_A; \mathbf{W}^T)) - R_{GT}^A(C_B, a_A) \right|^2 \\ & + \left| R^B(C_W, T_{AB}(a_A; \mathbf{W}^T)) - R_{GT}^A(C_W, a_A) \right|^2 \end{aligned} \right] \quad (5)$$

Where  $U(A_A)$  denotes the uniform distribution defined on the stroke control parameter space  $A_A$ , and the learning process of the translator is to minimize the objective function, i.e.

$$\min_{\mathbf{W}^T} J(\mathbf{W}^T) \quad (6)$$

Due to the differentiable property of the neural renderer  $R^B$ , this optimization problem is solved by stochastic gradient descent algorithm in this paper to complete the training of the stroke translator. For the sampled control parameter  $a_A$ , the corresponding neural renderer  $R_{GT}^A$  is first used to render the stroke  $R_{GT}^A(C_{W/B}, a_A)$  on a blank drawing board  $C_{B/W}$ , then the control parameter  $a_A$  is converted to  $a_B$  by the translator  $T_{AB}$ , and then the renderer  $R^B$  is used to obtain the rendering result  $R^B(C_{W/B}, T_{AB}(a_A; \mathbf{W}^T))$ , and finally calculate the  $L_2$  distance (i.e., MSE loss) of the two results to update the network parameters of the translator [27].

3) *Network structure:* Since the translation process of the two-stroke control parameters is equivalent to completing a kind of mapping, this paper adopts a four-layer fully-connected network to construct the stroke translator, the size of the input layer is the number of control parameters of the source stroke, and the size of the output layer is the number of control parameters of the target stroke.

## IV. DESIGN OF INTELLIGENT RENDERING GENERATION ALGORITHM BASED ON IMAGE TEXTURE

At present, although it is possible to render images, the processing of image details and textures is not satisfactory. In order to solve this problem, an intelligent painting generation algorithm based on image texture drawing technique is proposed and a hierarchical fusion generation countermeasure neural network is constructed. It renders the image using structural GAN and texture GAN, which further generates the texture details of the image and makes the image clearer and more realistic.

### A. Generative Adversarial Neural Network

Generative Adversarial Networks (GAN) is a new kind of deep generative mannequin proposed by using Goodfellow et al. It has been effectively utilized to picture generation, video generation, picture fashion migration and picture complementation and different scenarios.

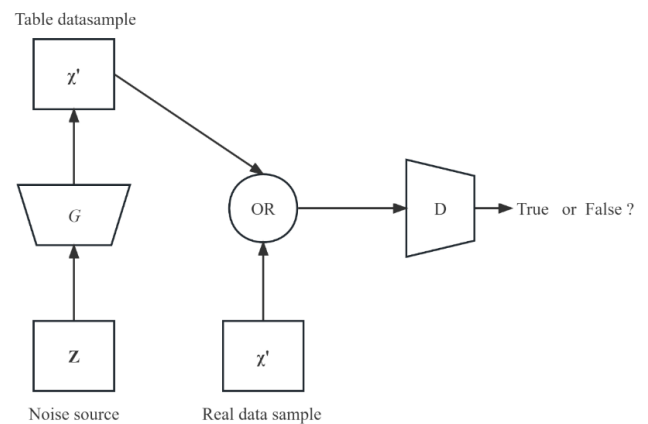


Fig. 8. Schematic diagram of GAN structure.

The generative adversarial community is stimulated via the zero-sum sport in recreation theory, which regards the trouble of producing sport statistics as a disagreement and sport between two networks, the discriminator and the generator. The position of the generator is to synthesize statistics in a given uniformly dispensed noise or generally dispensed noise [28, 29]. The two networks are trained, improved in the confrontation, and then continue the confrontation after repeated confrontations to obtain progress so that the generated data is constantly close to the real data until it cannot be distinguished from the real data so that the desired data content can be obtained. The principle of the GAN structure is shown in Fig. 8.

Definition G is a network to generate images, which can be called a generator, and z represents a random noise through the noise generated by the image can be recorded as G(z). D is a discriminative network, also known as a discriminator, whose role is to distinguish whether the noise generated by the image is real. If x represents an image, when the input parameter value x, the output D(x) represents the probability value that this image is a real image. When the value is equal to 1, it means that x must be a real image; and when the value is 0, it means that the image x is not a real image. During the training process, the generator and the discriminator play with each other; the mathematical relationship between the two can be expressed as follows:

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\ln D(x)] + E_{z \sim p_z(z)} [\ln(1 - D(G(z)))] \quad (7)$$

Where E is the probability function, x is the real data, z is the noise,  $p_{data}(x)$  denotes the distribution of the real data set,

D() denotes the discriminator,  $p_z(z)$  denotes the defined a priori noise, and G() denotes the generator.  $D(G(z))$  characterizes the probability of the D network to judge whether the image generated by G is real or not, and G, in order to make the image it generates converge as much as possible to the real image, needs to deceive the D network. G, in order to make the image generated by itself as close as possible to the real image, it needs to deceive the D network. Given any function D and G, there is a unique solution, and when G can generate a false image  $G(z)$ , the value of  $D(G(z))$  is always 0.5. As a result, the overall model can reach the global optimal state. In practice, it is found that maximizing  $\ln D(G(z))$  is better than minimizing  $\ln(1 - D(G(z)))$ .

Unlike typical coaching models, GAN implements two exclusive networks and an adversarial education method. It makes use of the returned propagation mechanism, in which a clearer and greater sensible pattern can be synthesized besides the complicated Markov Chain, and the computation is surprisingly simple Model Design

1) *Model architecture:* The paper proposes to generate an adversarial neural network based on the hierarchical fusion of image structure and texture to realize the intelligent generation of painting images so as to assist the creators to create better works. Its main model architecture is shown in Fig. 9

The sketches drawn are first fed into the system model for normalization preprocessing to ensure that the image conforms to the model.

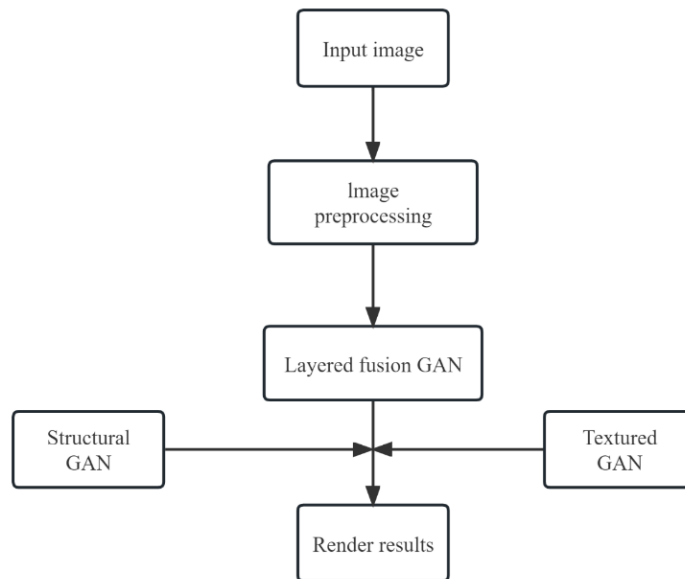


Fig. 9. The model architecture of the paper.

2) *Structural GAN:* The input of the generative network G is collected from uniformly distributed noise, and the output is

a structural line drawing. A 100-dimensional vector is used to represent the input noise z, and the size of the output image is

72×72×3. The discriminative network D will classify the generated structural line drawings based on the real images obtained from deep learning. The generative network adopts a 10-layer model and performs convolutional operations after passing through the fully connected layers to finally generate the structural line drawings. In the setup, batch normalization is used in each layer, and the ReLU activation function is used except for the last layer, which uses the Tanh activation function [30].

3) *Texture GAN*: The generated network is modified as a conditional GAN, that is, the conditional information is used as additional input for generator G and discriminator D. As an addition to discriminator D, the structure line map not only makes the generated image more real after input, but also requires the generated image to match the structure line map to ensure its controllability. When training the discriminator, only the real image and its corresponding structural line drawings are considered as positive examples, so that a higher resolution 128×128×3 image can be generated by texture GAN.

For example, if the image  $x=(x_1, \dots, x_M)$ , its corresponding structural line drawing is  $C=(C_1, \dots, C_M)$ , and the uniformly distributed noise is  $\hat{z}=(\hat{z}_1, \dots, \hat{z}_M)$ , the generating function can be changed from  $G(\hat{z}_i)$  to  $G(C_i, \hat{z}_i)$ , and the discriminative function can be changed from  $D(x_i)$  to  $D(C_i, x_i)$  losses of the discriminative network can be rewritten as follows, respectively.

$$L_{\text{cond}}^G(C, \hat{z}) = \sum_{i=M/2+1}^M L(D(C_i, G(C_i, \hat{z}_i)), 1) \quad (8)$$

$$L_{\text{cond}}^D(X, C, \hat{z}) = \sum_{i=1}^{M/2} L(D(C_i, x_i), 1) + \sum_{i=M/2+1}^M L(D(C_i, G(C_i, \hat{z}_i)), 0) \quad (9)$$

In the generator network architecture, a 128×128×3 textured line graph and 100-dimensional vector noise  $\hat{z}$  are taken as inputs to the network. The inputs to the network are the structured lines and the concatenation of images. The two will first go into the convolutional layer and the inverse convolutional layer, respectively, and then merge to form the 32×32×192 feature maps and further perform seven layers of convolution and inverse convolution on the top of these feature maps. The final output is a 128×128×3 image.

4) *Joint model*: The discriminator in the texture GAN treats the generated pattern line drawings and images as negative samples and the real structure line drawings and real images as positive samples [31]. As a result, the generator network loss function of the structural GAN can be expressed as follows:

$$L_{\text{joint}}^G(\hat{z}, \tilde{z}) = L^G(\hat{z}) + \lambda \cdot L_{\text{cond}}^G(G(\hat{z}), \tilde{z}) \quad (10)$$

Where  $\hat{z}$  and  $\tilde{z}$  represent two sets of uniformly distributed samples in structural GAN and texture GAN, respectively. The first term on the right side of the equation represents the adversarial loss of the structural GAN discriminator, and the second term on the right side of the equation represents the loss of the texture GAN.  $\lambda$  is a hyperparameter, and its value is set to 0.1 in this experiment, which is smaller than the learning rate of the texture GAN, so as to avoid the occurrence of the overfitting situation.

## B. Test Results and Analysis

1) *Experimental environment*: The experimental environment is Python 3.6, the processor is i7-6800k, the memory is 32 GB, the graphics card is GTX2080Ti, the operating system is Linux, and the experiment is based on the open-source deep learning framework PyTorch for simulation [32].

The experiments were optimized using Adam optimizer, where the momentum term  $\beta=0.5$ ,  $\beta_2=0.999$ , and the batch size M is 128, and the inputs and outputs of all the networks are scaled to [-1, 1]. The learning rate of both structural GAN and texture GAN is set to 0.0002. In joint learning, the learning rate of texture GAN is set to  $10^{-6}$ , and the learning rate of structural GAN is set to  $10^{-7}$ . In addition, the rest of the parameter settings are designed according to the parameters of the experiments in DCGANI.

2) *Experimental parameter settings*: All algorithms involving stochastic gradient descent in this paper were trained using the Adam optimizer. The neural renderer and translator learning rates are updated using an exponential update strategy every 100 Epochs, and the reinforcement learning uses a manual method to update the learning rates. The dimensions of both the drawing board and the target image are set to  $H \times W \times 3 = 128 \times 128 \times 3$ . Noting that the limited number of steps for the drawing intelligences is 40 and the size of the action combination is 5, which corresponds to a total number of strokes of  $40 \times 5 = 200$ .

3) *Experimental results*: It can be found that after the fusion of the generative adversarial neural network, the final image is more realistic in texture and does not change the original object that the painter is trying to depict, so the model works well.

In Fig. 10, as can be seen from the figure, according to the training method proposed in this paper, the model can converge to a stable value within 140 Epochs, which indicates the feasibility of the training method proposed in this paper. For different strokes, the neural renderer converges to different values: the point-like texture of chalk strokes is more complex, which is difficult for the neural renderer to learn. Watercolor strokes and oil strokes are relatively simple, and thus converge to higher precision, but due to the difficulty in learning the texture of the details of the oil strokes, the convergence precision is relatively low. This may be due to the fact that the strokes are relatively “light,” and thus, the absolute value of the noise is small, resulting in a high PSNR value.



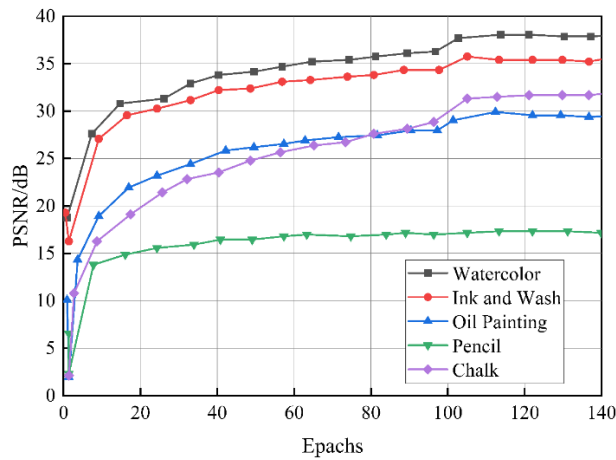


Fig. 10. Stroke renderer training content.

The translator is further applied to the reinforcement learning drawing simulation framework to translate the stroke control parameters generated by the intelligent body, and the drawing results of different stroke styles can be obtained.

Overall, the translated drawing results can basically maintain the content of the original results, but the details may be lost, and the differences between different strokes are huge, as shown in Fig. 11.

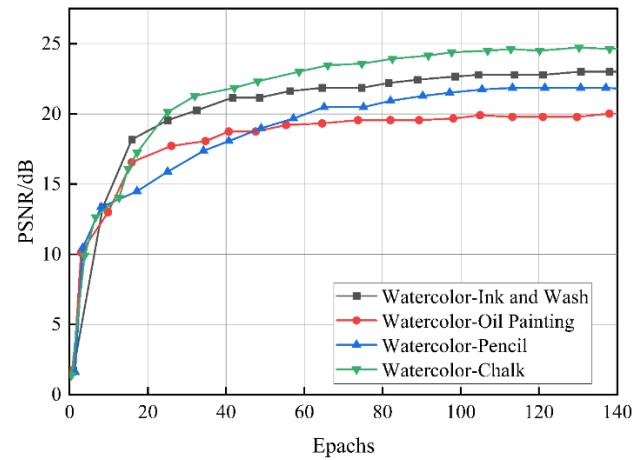


Fig. 11. Stroke translator training process.

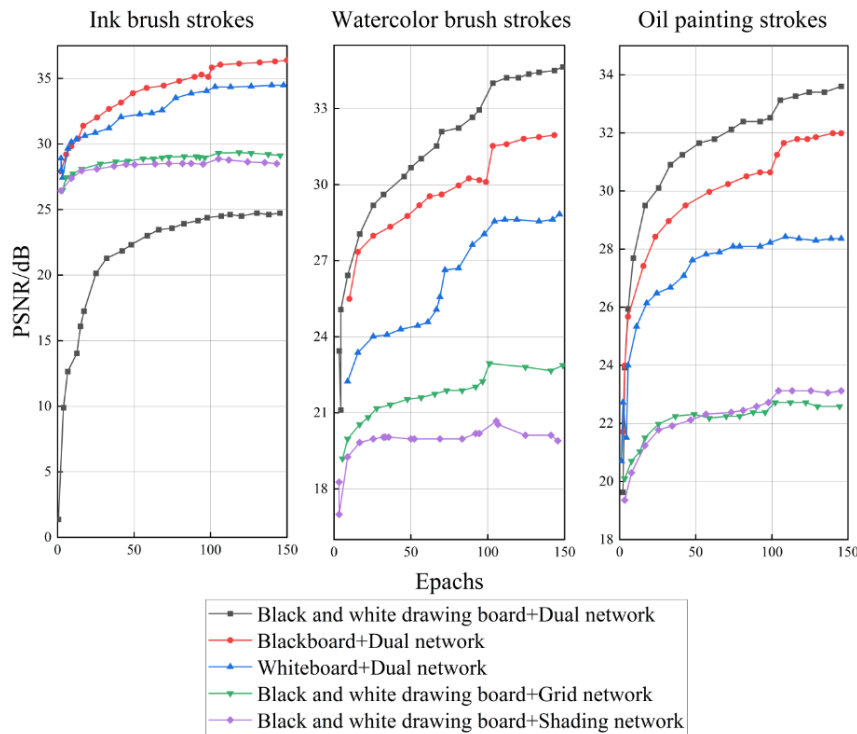


Fig. 12. Neural renderer ablation experiments.

In the case of training using black and white panels, neither the rasterization network nor the coloring network alone can achieve the best results, illustrating the effectiveness of the two-way network. In the case of using the two-way network, neither the black drawing board alone nor the white drawing board alone can achieve the best results, illustrating the effectiveness of the training method proposed in this paper (as shown in Fig. 12).

4) *Experimental analysis and comparison*: The paper also quantitatively analyzes the generated images and evaluates the

quality of the generated images by the international common evaluation index IS. The larger the value of IS, the higher the quality of the generated images and the more realistic the effect, which mainly examines two aspects of the image performance: 1) whether the generated images are clear or not; 2) whether the generated images are varied or not.

In GAN, the conditional probability  $p(y|x)$  is usually expected to be highly predictable. Where  $x$  denotes a given image, and  $y$  denotes a specific object contained in the image.

IS uses a fixed classification network, Inception Network, to realize the classification of the generated image and then predicts  $p(y|x)$ . In terms of diversity, the edge probabilities are computed using the following formula.

$$\int_z p(y|x = G(z)) dz \quad (11)$$

In summary, the formula for IS is as follows:

$$IS(G) = e^{E_{x \sim p} D_{kl}(p(y|\square) \| p(y))} \quad (12)$$

where,  $D_x$  is the discriminant function containing the KL-Divergence constraint. In order to better validate the effect of this experiment, 1000 images were also generated using DCGAN and ProGAN respectively, and 10 evaluations were done using MNIST samples to take the mean value. The experimental results obtained are shown in Table II.

TABLE II. IS ASSESSMENT RESULTS

arithmetic	IS
GAN	0.8
DCGAN	1.44
ProGAN	1.87
Methodology of this paper	2.30

As can be viewed from Table I, the IS fee of the paper's approach is the highest, which shows that the photo generated the usage of the hierarchical fusion facets is of greater nice and has clearer and extra practical important points.

In this section, the portray consequences generated by way of the techniques in this paper are in contrast (the goal photograph is from the take a look at dataset), the quantitative assessment is completed via calculating the similarity index between the portray consequences generated by way of every approach and the goal photograph (taking the common price of the complete dataset), and the qualitative evaluation is finished through examining the true portray effects of one of a kind methods.

The parameters of several comparison methods were set as follows:

Using ink strokes, in order to make the size of the output image 128×128, the grid is set to 3×3, 23 strokes per grid totaling 207 strokes, and the number of iteration steps is 50;

Using watercolor brushstrokes, the size of the board is 128×128; using the pre-training model and parameters provided in the original article, the number of painting steps is 40, corresponding to the number of strokes is 40×5=200;

Using oil brush strokes, the board size is 128×128, and the number of strokes is limited to 200.

To summarize the floor plan rendering capabilities of the model: 1) the quality of CyceGAN rendered floor plan meets the need for clear presentation; 2) Epoch50 rendering is flatter, Epoch300 is rich in detail but has a certain chance of producing pixelated details, 3) the rendering of the nature of the land is accurately expressed, but the rendering of the

structures is not prominent enough and needs to be manually further emphasized, and 4) the middle of the rendering colors is not as good as the middle of the rendering, and there is a certain lack of transition tones which are somewhat missing.

## V. CONCLUSION

With the rapid development of Web technology, basic application systems based on b/s architecture have been widely developed and applied, and software based on b/s architecture has become the trend of software development. As a common form of artistic creation, painting is an important means of visual communication in the fields of news dissemination, prototype design, movie and television creation. The content of digital painting is easier to disseminate in the current network era and has important research and application value. In this paper, based on the previous research work, an image generation algorithm is proposed based on the intelligent rendering algorithm of image texture, which generates the structural line drawing through the structural GAN, and then through the rendering of the texture GAN to generate the ideal effect of the painting. The main conclusions are:

1) *This* chapter mainly introduces the related technical theories involved in the art image processing platform, the B/S network architecture which is popular among developers nowadays, the Python programming technology which has a great momentum, and the Django framework which is convenient for researchers in Python technology.

2) *This* chapter first introduces the model-based DDPG reinforcement learning approach as the main framework and clarifies in detail how reinforcement learning is applied to the task of painting simulation through three aspects: modeling, learning algorithm, and network architecture. Then, we introduce a two-way neural renderer for constructing the drawing simulation environment and a stroke translator that can realize the conversion of stroke control parameters.

3) *Based* on the existing open-source drawing board and renderer technologies, this paper defines a variety of stroke rendering methods and integrates them into a unified renderer framework to construct various drawing simulation environments. On this basis, this paper uses a unified two-way neural network structure and training method to realize a neural renderer that can be used for different stroke rendering. From the experimental results, it can be seen that the two-way neural renderer used in this paper combined with the DDPG reinforcement learning framework can effectively generate the drawing content of multiple stroke styles and realize the simulation of the drawing process; the stroke translator proposed in this paper can complete the expected translation of stroke control parameters and realize the migration of multiple stroke drawing styles.

## REFERENCES

- [1] Jiang Y. Computer Vision Object Detection Algorithm based on Convolutional Neural Network. Journal of Shenyang University of Technology, 2021, 43(5): 557-562.
- [2] Chen J, Du M, Zheng J, et al. Double Level stroke line simplification

- Method based on drawing time sequence. Journal of Computer-Aided Design and Graphics, 2019, 9.
- [3] Kong S, Yin J. Design of real-time rendering system for 3D animation image texture. Modern Electronic Technique, 2018, 41(5): 102-105.
- [4] Xu Q, Zhong S, Chen K, et al. Optimal selection method of CycleGAN Cycle consistent loss coefficient in image generation with different texture complexity. Computer Science, 2019, 46(1): 100-106.
- [5] Ma Y, Xu X, Zhang R, et al. Research progress of generative adversarial networks and their applications in image generation. Journal of Computer Science and Exploration, 2021, 15(10): 1795.
- [6] Chen F, Zhu F, Wu Q, et al. Review of generative adversarial networks and their applications in image generation. Chinese Journal of Computers, 2021, 44(2): 347-369.
- [7] Liang J, Wei J, Jiang Z. A review of generative adversarial networks. Exploration of Computer Science and Technology, 2020, 14(1): 1-17.
- [8] Wang K, Gou C, Duan Y, et al. Research progress and prospects of generative adversarial networks. Acta Automatica Sinica, 2017, 43(3): 321-332.
- [9] Ye W, Gao H, Weng S, et al. A two-stage art font rendering method based on CGAN network. Journal of Guangdong University of Technology, 2019, 36(03): 47-55.
- [10] Zhu Danni, Xu Xiao-Hua, He Jing-Jing, et al. Non-reference super resolution image quality evaluation using multi-layer perceptron regression. Journal of Xi'an Polytechnic University, 2022, 36(5).
- [11] Popa T, Ibanez L, Levy E, et al. Tumor volume measurement and volume measurement comparison plug-ins for VolView using ITK. Medical Imaging 2006: Visualization, Image-Guided Procedures, and Display. SPIE, 2006, 6141: 395-402.
- [12] Tang C, Yuan J, Xia C. Research and design of medical image reading system based on B/S structure. Computer and Digital Engineering, 2014, 42(2): 311-314.
- [13] He J, Bao Y, Zhang J, et al. Design and implementation of Medical Laboratory Information Platform (LIS) based on B/s model. Computer Application and Software, 2016, 33(3): 83-86.
- [14] Mukai Y U, Tokoi K. Watercolor Rendering with Consideration of Motion.
- [15] Frans K, Soros L, Witkowski O. Clipdraw: Exploring text-to-drawing synthesis through language-image encoders. Advances in Neural Information Processing Systems, 2022, 35: 5207-5218.
- [16] Huang X, Belongie S. Arbitrary style transfer in real-time with adaptive instance normalization. Proceedings of the IEEE international conference on computer vision. 2017: 1501-1510.
- [17] Kotovenko D, Sanakoyeu A, Ma P, et al. A content transformation block for image style transfer. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 10032-10041.
- [18] Lee J, Hwangbo J, Wellhausen L, et al. Learning quadrupedal locomotion over challenging terrain. Science Robotics, 2020, 5(47): eabc5986.
- [19] Li G. Application of 3ds Max Rendering Technology in Virtual Scene Design Experimental Teaching. Art Education Research, 2015 (1): 170-173.
- [20] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 2015, 28.
- [21] Wang T, Gao X. A rendering method of multi-stroke anisotropic Van Gogh style oil painting. Computational Technology and Automation, 2017, 36(2): 125-128.
- [22] Sanakoyeu A, Kotovenko D, Lang S, et al. A style-aware content loss for real-time hd style transfer. proceedings of the European conference on computer vision (ECCV). 2018: 698-714.
- [23] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. Proceedings of the IEEE international conference on computer vision. 2017: 2223-2232.
- [24] Kim T, Cha M, Kim H, et al. Learning to discover cross-domain relations with generative adversarial networks. International conference on machine learning. PMLR, 2017: 1857-1865.
- [25] Levine S, Pastor P, Krizhevsky A, et al. Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection. The International journal of robotics research, 2018, 37(4-5): 421-436.
- [26] Zhang L, Ping X, Zhang T. Image information camouflage algorithm with first-order statistical feature preserving. Journal of Computer-Aided Design and Graphics, 2005, 17(1): 99-104.
- [27] Guan Q, Zhu J, Zhao X, et al. An image steganography method based on linear programming feature selection and integrated classifier. Journal of Cyber Security, 2018, 3(1).
- [28] Zhang R, Isola P, Efros A A, et al. The unreasonable effectiveness of deep features as a perceptual metric. Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 586-595.
- [29] Isola P, Zhu J Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1125-1134.
- [30] Sekhon A. Synthesizing Programs for Images using Reinforced Adversarial Learning.
- [31] Kosugi S, Yamasaki T. Unpaired image enhancement featuring reinforcement-learning-controlled image editing software. Proceedings of the AAAI conference on artificial intelligence. 2020, 34(7): 11296-11303.
- [32] Silver D, Huang A, Maddison C J, et al. Mastering the game of Go with deep neural networks and tree search. nature, 2016, 529(7587): 484-489.

# Using EEG Effective Connectivity Based on Granger Causality and Directed Transfer Function for Emotion Recognition

Weisong Wang, Wenjing Sun\*

School of Marxism, XinJiang Normal University, WuLuMuQi 830017, Xinjiang, China  
School of Marxism, Guangxi Science and Technology Normal University, LaiBin 546199, Guangxi, China

**Abstract**—Emotion is a complex phenomenon that originates from everyday issues and has significant effects on individual decisions. Electroencephalography (EEG) is one of the widely used tools in examining the neural correlates of emotions. In this research, two concepts of Granger causality and directional transfer function were utilized to analyze EEG data recorded from 36 healthy volunteers in positive, negative and neutral emotional states and determine the effective connectivity between different brain sources (obtained through independent component analysis). Shannon entropy was utilized to sort the brain sources obtained by the ICA method, and average topography helps to add spatial information to the proposed connectivity models. According to the obtained confusion matrix, our method yielded an overall accuracy of 75% in recognizing three emotional states. Positive emotion was recognized with the highest accuracy of 87.96% (precision = 0.78, recall = 0.78 and F1-score = 0.81), followed by neutral (accuracy = 82.41%) and negative (accuracy = 79.63%) emotions. Indeed, our proposed method achieved the highest recognition accuracy for positive emotion. The proposed model in the present study has the ability to identify emotions in a completely personalized way based on neurobiological data. In the future, the proposed approach in the present study can be integrated with machine learning and neural network methods.

**Keywords**—EEG; effective connectivity; granger causality; directed transfer function; emotion recognition

## I. INTRODUCTION

Emotion is a complex phenomenon that originates from everyday issues and has significant effects on individual decisions. The extent of these decisions can affect the personal and social life of the people of society [1]. Emotions are very important in learning and communicating, and their expression plays a big role in human relationships. Emotions can directly affect a person's performance, and therefore, it is important to try to understand their sources and control them [2]. A certain emotion is one of the brain states that is produced by the electrical activity of millions of neurons and is associated with physiological changes in the whole body [3]. When experiencing emotions, the brain's left and right hemispheres are involved in different ways [4]. For example, it has been shown that the left prefrontal area is more involved in dependent emotional reactions and the right prefrontal area is involved in withdrawal reactions [5]. Generally, it is hypothesized that the left hemisphere is dominant in positive

emotions, and the right hemisphere is dominant in negative emotions [6].

Electroencephalography (EEG) is one of the widely used tools in examining the neural correlates of emotions [7]. This instrument has many applications in cognitive neuroscience and psychiatry and improves our insight into various behavioral phenomena including depression [8], bipolar disorder [9-17] and hyperactivity [18-22]. Many researchers have applied various machine learning techniques to EEG data to develop an automated EEG-based emotion recognition system [23]. Zhang et al. applied the Empirical Mode Decomposition (EMD) technique to EEGs, calculated the sample entropy from the first four IMFs, and reported an average accuracy of 93.20% in classifying five distinct emotions [24]. Zheng and Lou trained deep belief networks by calculating entropy features from different EEG frequency bands and achieved an accuracy of 86.65% in detecting three negative, neutral and positive emotions [25]. Meng et al. proposed an EEG-based emotion recognition system based on entropy features and cascaded convolutional recurrent neural networks and achieved an accuracy of 94.85% in valence-based classification problems [26]. Hwang et al. utilized convolutional neural networks along with generating topology-keeping differential entropy features to prevent the loss of localized information and achieved an average accuracy of 90% in recognizing three negative, positive and neutral emotions [27]. Nawaz et al. proposed an EEG-based emotion recognition system based on various linear and nonlinear features such as fractal dimension and wavelet energy, feature selection by principle component analysis, and multiple classifiers such as SVM and KNN, and reported an average accuracy of 77.60%, 78.96% and 77.62% in detecting dominance, arousal and valence emotions, respectively [28]. Two recent reviews on EEG-based emotion recognition emphasized the importance of such systems for various scientific fields, such as psychology and cognitive neuroscience, and highlighted the necessity of developing these systems with a special focus on regional brain connectivity [29, 30].

In the last two decades, neuroscience studies have focused on the connections of different areas of the cerebral cortex and how these areas interact when performing a specific sensory-motor or cognitive action to better understand brain function [31]. Many attempts have been made to quantify these connections through EEG analysis [32]. Functional

\*Corresponding Author, e-mail: sxiaojing202305@163.com

connectivity is an observable phenomenon that can be quantified by measuring statistical dependencies such as correlation or transfer entropy [33]. On the other hand, effective connectivity refers to the model parameters that attempt to explain the observed dependencies (functional connectivity). From this point of view, effective connectivity refers to the concept of coupling or direct causal influence by examining the direction of information dissemination [34, 35]. Bagherzadeh et al. achieved a high accuracy of 99% in the classification of five emotional states using effective EEG connectivity and convolutional neural networks [36]. In another study, Bagherzadeh et al. used EEG frequency effective connectivity maps based on the transfer learning technique and achieved an accuracy of 95% in the classification of five emotional states [37]. Although these studies have shown the high potential of EEG effective connectivity as a biomarker for human emotion recognition, few studies have used this important feature of brain signals to develop EEG-based emotion recognition systems.

Therefore, in this research, we aim to process EEG signals to estimate effective connectivity in brain resources and provide models to investigate how brain resources influence each other in order to detect three emotional states: neutral, negative, and positive. For this purpose, we investigated the potential of Granger causality and directed transfer function to estimate effective cortical connectivity in different emotions. Our research showed that these approaches can reveal noticeable effective connectivity patterns in each emotion, which can be classified in order to emotion recognition.

## II. METHODS

Fig. 1 shows the work process of the current research. Considering that in this research, two concepts of Granger causality and directional transfer function are used in EEG data processing, it is necessary to briefly introduce these concepts first.

### A. Granger Causality

Granger causality is a technique to derive specific kinds of causal dependence among stochastic samples by reducing the

bias of predicting possible effects if past observations of the hypothesized causes are utilized to anticipate the effects plus previous observations of the possible effects. This concept was first proposed through Wiener and then reformulated through Granger based on linear autoregressive models. This algorithm considers two assumptions: (I) a cause should precede its effects, and (II) data on the past of a cause should enhance the anticipation of the effects above and beyond data on the collective past of the other observed samples. To estimate influences from the channel  $x_j$  to  $x_i$  for  $n$  channel autoregressive processes, we considered  $n$  and  $n-1$  multivariate autoregressive models. The model is fitted to overall  $n$ -channel system, resulting in the residual variance  $V_{i,n}(t) = \text{var}(E_{i,n}(t))$  for time series  $x_i$ . Then, a  $n-1$  multivariate model is fitted for  $n-1$  channels, except for channel  $j$ , which results in the residual variance  $V_{i,n-1}(t) = \text{var}(E_{i,n-1}(t))$ . So, Granger causality is given by:

$$Granger_{j \rightarrow i}(t) = \ln \left( \frac{V_{i,n}(t)}{V_{i,n-1}(t)} \right) \quad (1)$$

### B. Directed Transfer Function

This method is used to measure the direction and frequency content of brain activities. The directed transfer function is a multivariate approach that is formulated by multivariate autoregressive models to investigate the causal relationships between EEG channels and recognize the directed dissemination of EEG activities. This algorithm works in the frequency domain to characterize regional connectivity based on the factorization of the coherence between two EEG channels.

### C. Dataset

In this research, we utilized an EEG dataset on neutral, positive, and negative emotional states from [38]. This dataset includes EEG signals recorded from 36 young adults in the neutral, positive, and negative emotional states induced by standard images from the International Affective Picture System (IAPS). 16 Ag/AgCl electrodes were utilized to capture brain signals at a sampling rate of 512 Hz based on a 10-20 international recording protocol.

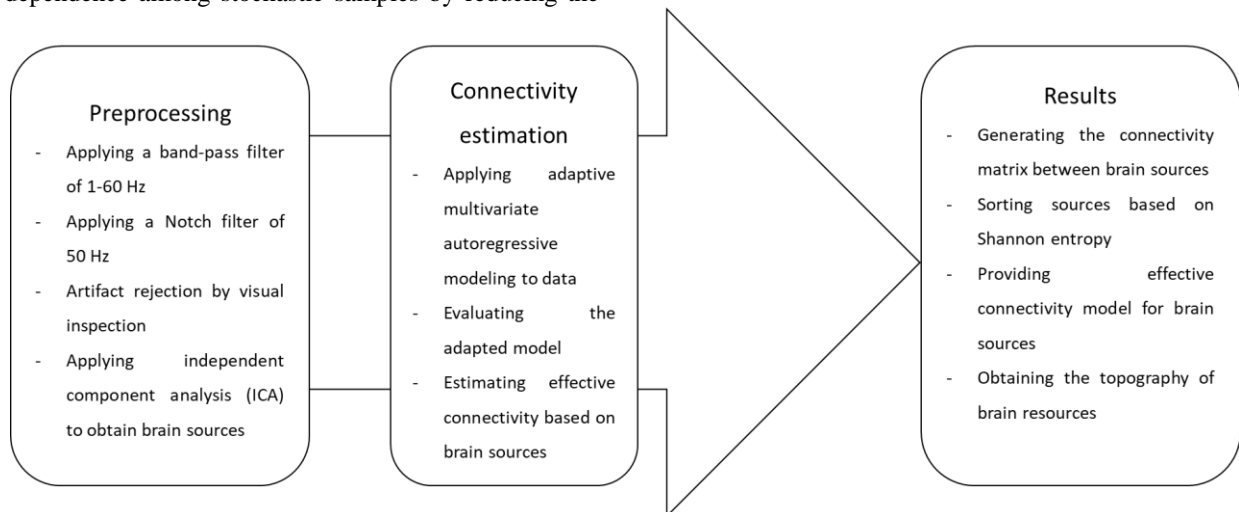


Fig. 1. Flow chart of the proposed method for EEG-based emotion recognition based on effective connectivity.

#### D. Data Preprocessing

First, we applied a Butterworth band-pass filter (order 4) with a frequency of 1–60 Hz to the signals to cancel unwanted noise. A 50 Hz notch filter was then applied to the signals to cancel power line interference. Next, an experienced neurologist checked all EEGs and rejected all parts of the signals contaminated with various artifacts, such as body motion and eye blinking. After obtaining a clean EEG signal for each subject, independent component analysis (ICA) was applied to the signals to decompose them into independent components. A fast ICA algorithm was utilized in this work to decompose 16 source components from EEG channels. However, we encountered a bad situation when using ICA because the order of the extracted components changes in every run of the algorithm. To solve this problem, Shannon's entropy of sources was used as a measure of information to sort brain sources. Sources are sorted according to Shannon entropy value and in ascending order. These brain sources obtained from ICA were used for further analysis.

#### E. Effective Connectivity Estimation

EEGLAB and SIFT toolbox were utilized to estimate effective connectivity [39]. The first step to estimate the connectivity value is to fit a model to the data using Adaptive Multivariate Autoregressive (AMVAR) modeling, which is acceptable for this purpose. To adapt this model, the length of the window, the step of the window, and the order of the model should be selected. After selecting the values of these parameters, the accuracy of the fitted model is evaluated. Table I shows the selected values for these parameters.

TABLE I. SELECTED VALUES FOR AMVAR MODEL PARAMETERS

Parameter	Value
Window length (second)	5
Window step (second)	1
Model order	10

It should be noted that the higher order of the model causes complexity, and the short length of the window leads to an increase in calculations. As a result, a trade-off should be made in choosing these parameters. If the model is sufficiently adapted to the data, the residual coefficients of the model should be small and uncorrelated relative to the real data. The presence of correlation in the residuals indicates the presence of correlated structures in the data that the model is unable to provide. To solve this issue, a null hypothesis with a significance level is considered, and the model is evaluated through it. To evaluate the whiteness of the residuals in this research, the autocorrelation function method with a significance level of 90% was utilized, as well as the consistency assessment through the percent consistency [40]. The stability of the model was evaluated with the stability index. A VAR model is stable if the augmented coefficient matrix of all stability indices is less than one [41].

Granger causality and direct transfer function methods were used to estimate the effective connectivity. In this study, effective connectivity was estimated in four EEG frequency bands: delta (1-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), and beta

(13-30 Hz). In fact, for six sources sorted by Shannon entropy, effective connectivity was estimated in these frequency bands.

#### F. Performance Evaluation

To evaluate our proposed approach for emotion recognition, we utilized various classification metrics, including accuracy, precision, recall and F1-score. These metrics are given by:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (5)$$

where TP, TN, FP and FN denote true positive, true negative, false positive and false negative elements for a specified emotional class obtained from the confusion matrix.

### III. RESULTS

First, the accuracy, stability, and consistency of the model were checked using the mentioned criteria. The results showed that the stability of all models was fully established, and the average model stability in all samples was about 78%. The average rejection of the null hypothesis was 93%. For each emotional state, the maximum number of connections in each frequency band was calculated. In each sample, a source can receive information from other sources. Fig. 2 shows a schematic of the calculation of the connectivity between sources as a sender or receiver of information.

Based on the schematic shown in Fig. 2, the sources with the most receiving and sending information in three emotional states were determined based on averaging in different frequency bands as shown in Table II.

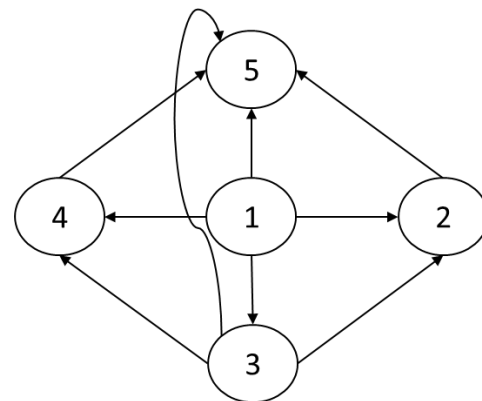


Fig. 2. A schematic of the calculation of the connectivity between sources as a sender or receiver of information. Source 1 is the sender with 100%, and source 5 is the receiver with 100%.

Based on the results obtained in Table II, the proposed model based on Granger causality and directed transfer function for effective connectivity for neutral, negative, and positive emotional states is shown in Fig. 3. The models presented in Fig. 3 are obtained from the statistical analysis of the connection between different sources. In fact, in these

models, the information related to the location of brain sources is not included, so they cannot be used to identify the location of connections. To solve this problem, topographic images of sources were used.

TABLE II. SOURCES WITH THE MOST RECEIVING AND SENDING INFORMATION IN THREE EMOTIONAL STATES

Emotional state	Sending sources		Receiving sources	
	Source	Percentage	Source	Percentage
Neutral	5	54.2	9	52.1
	7	48.4	7	46.8
Negative	6	47.1	8	48.7
	10	44.3	6	44.5
Positive	10	66.2	6	66.2
	9	43.1	7	49.2

In this model, each circle represents a brain source, which is assigned a number based on Shannon entropy ranking.

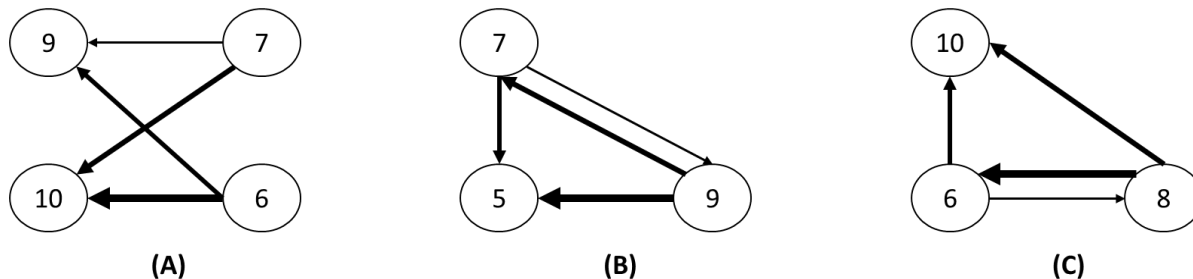


Fig. 3. Proposed model based on Granger causality and directed transfer function for effective connectivity for (A) positive, (B) neutral, and (C) negative emotional states.

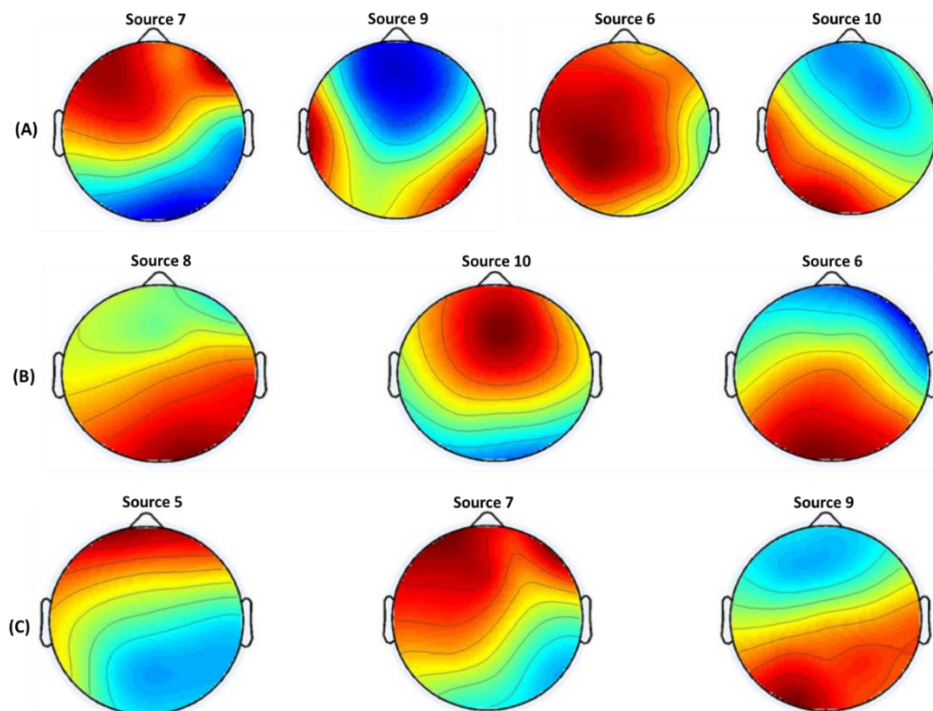


Fig. 4. Average topography images of 36 samples for sources with the most input and output connections, as specified in Fig. 3, extracted with independent component analysis, in (A) positive, (B) negative, and (C) neutral emotional states.

Arrows go from the source of the information transmitter to the source of the information receiver. The larger the diameter of an arrow, the greater the amount of information spread.

In addition to separating different sources from each other, ICA provides us with information about the location of these sources on the scalp by producing images called the topography of a source. These images are important in that they can show how a source spatially affects the entire scalp. As a result, by considering the assumptions about the location that is activated during specific stimuli, the location of their sources can be identified. Accordingly, the topography of the sources whose models were created in the previous section was determined, and their average topography was examined in all samples. It is worth mentioning that the selected sources are the same in terms of entropy. The topography of source 10 in the positive emotion belongs to source 10 in terms of entropy in all samples. Table II and Fig. 3 specify the sources with the most output and input connections in each emotion. Based on this, Fig. 4 shows the average topography of these sources for positive, negative, and neutral emotional states.

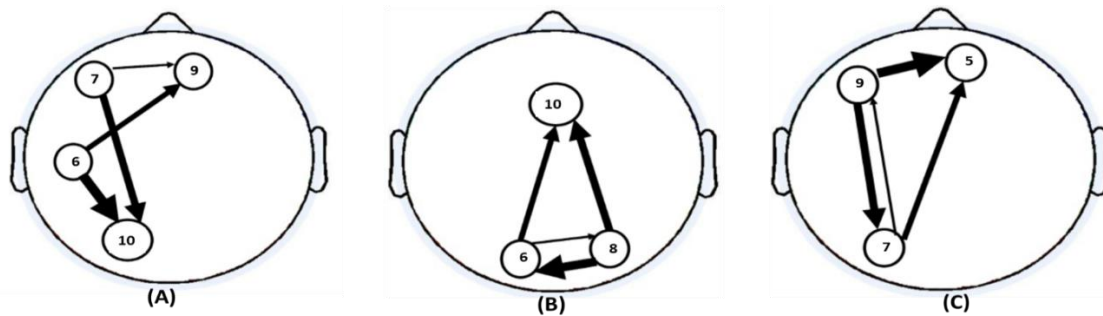


Fig. 5. The proposed models for effective connectivity in (A) positive, (B) negative, and (C) neutral emotional states with spatial information.

The images in Fig. 4 contain spatial information about brain sources. By combining these images with the information of the models from Fig. 3, other new models are presented that also contain spatial information about brain sources. Fig. 5 shows the presented model for effective connectivity between brain sources in positive, negative, and neutral emotional states with spatial information.

In this model, each circle represents a brain source, which is assigned a number based on Shannon entropy ranking. Arrows go from the source of the information transmitter to the source of the information receiver. The larger the diameter of an arrow, the greater the amount of information spread.

The models presented in Fig. 5 are the average of all samples. In the next step, similarity measurement and labeling operations were performed for each sample. In other words, each individual's connectivity model was compared with the average models presented in Fig. 5 and labeled based on its similarity to each of the average models. The total number of samples is 36 subjects with three emotional states: positive, negative, and neutral. Therefore, there are 108 models with similar distribution in different emotional states. To measure the similarity of each sample in all emotional states and frequency bands, first, the average input and output information of all sources of that sample was calculated. The rate of receiving and sending information from all sources was determined for each sample. Then, in each sample, two sources that had the highest amount of receiving and sending information were selected. In the next step, the emotional state label of 108 samples was removed, and the samples were shuffled. Two selected sources from each sample were compared with the proposed models and given a label according to their similarity to each of the models as shown in Table III.

TABLE III. THE RESULT OF LABELING SAMPLES FOR POSITIVE, NEGATIVE, AND NEUTRAL EMOTIONAL STATES IS BASED ON THE PROPOSED EFFECTIVE CONNECTIVITY MODELS

		Estimated label based on similarity with proposed models		
		Positive	Neutral	Negative
Real label	Positive	28	3	5
	Neutral	2	26	8
	Negative	3	6	27

Table III shows the confusion matrix obtained by the proposed approach based on effective connectivity to identify positive, negative, and neutral emotions. According to the

obtained confusion matrix, our method yielded an overall accuracy of 75% in recognizing three emotional states. Table IV shows the obtained classification criteria for each class. As shown, positive emotion was recognized with the highest accuracy of 87.96% (precision = 0.78, recall = 0.78, and F1-score = 0.81), followed by neutral (accuracy = 82.41%) and negative (accuracy = 79.63%) emotions.

TABLE IV. CALCULATED CLASSIFICATION CRITERIA FOR EACH POSITIVE, NEGATIVE, AND NEUTRAL EMOTIONAL STATE

Class	n (truth)	Accuracy (%)	Precision	Recall	F1-score
Positive	33	87.96	0.78	0.85	0.81
Neutral	35	82.41	0.72	0.74	0.73
Negative	40	79.63	0.75	0.68	0.71

#### IV. DISCUSSION

The aim of this study was to estimate cortical effective connectivity from the EEG signal of emotions based on Granger causality and directed transfer function for the recognition of different human emotions. Our proposed method achieved the highest recognition accuracy for positive emotion. In the proposed model for positive emotion, source 10 had the highest amount of receiving information from other sources. This finding is consistent [38], where source 10 achieves the highest recognition accuracy for positive emotion. In the proposed model for negative emotion, source 8 had the highest amount of sending information to other sources. Abdolssalehi et al. also achieved the highest recognition accuracy for negative emotion through source 8 [38]. However, in [38], the authors relied on recurrence quantification analysis for pattern recognition from brain sources.

In contrast, we utilized the effective connectivity between these sources and improved the results obtained in Abdolssalehi et al.'s work. In addition, as mentioned, sources 10 and 8 had the highest amount of information exchange in positive and negative emotional states, respectively. The average topography of sources 10 and 8 of all samples was located in the left posterior hemisphere and the right posterior hemisphere, respectively. This finding is consistent with the valence hypothesis, which assumes the opposite predominance of the right hemisphere for negative emotions and the left hemisphere for positive emotions [49, 50]. Table V compares the findings of the present study with previous studies. As can be seen, our proposed system performs better than most of the previous methods.



TABLE V. COMPARING THE FINDINGS OF THE PRESENT STUDY WITH PREVIOUS STUDIES

Reference	Algorithm	Results
[42]	Spectral features and SVM classifier	49.4% like, 55.7% arousal, 58.5% valence
[43]	Spectral features and SVM classifier	79.59% sadness, 74.11% anger, 86.15% joy, 83.59% pleasure
[44]	Spectral and statistical features and LDA classifier	62% anxiety, 50% engagement, 57% boredom
[45]	Spectral assymetry index, wavelet entropy and SVM classifier	82.5% negative excitement, 64% neutral
[46]	Spectral features, Gabor transform and probabilistic neural network	62.97% sadness, 69.74% disgust, 73.64% anger, 56.79% fear
[47]	Spectral features and SVM classifier	50.5% valence, 62.1% arousal
[48]	Spectral features and SVM classifier	61% sadness, 58% fear, 53% anger, 51% joy
Our proposed approach	Effective connectivity and confusion matrix	87.96% positive, 82.41% neutral, 79.63% negative

Most of the studies that have tried to recognize human emotions using EEG signals have used a variety of machine learning methods and artificial neural networks that provide us with a black box-like function [51-53]. Although some of these studies have reported very high detection accuracies, the process of pattern recognition in them is unclear and ambiguous [54]. However, in this research, we tried to choose and propose a transparent work process to overcome this important limitation of previous studies. Although we did not use learning machines and artificial neural networks for classification and pattern recognition, the obtained results are quite promising. Therefore, our proposed model can be very useful in various research fields, such as psychiatry, psychology, neuroscience, and cognitive science. The important thing to consider about our proposed model is that this model can work completely depending on a person, and it is a personalized model. Individual and cultural differences are very important issues in the development of emotion recognition systems. At the same time, most previous techniques ignore this important issue. However, the proposed model in the present study has the ability to identify emotions in a completely personalized way based on neurobiological data. In the future, the proposed approach in the present study can be integrated with machine learning and neural network methods and improve the proposed model.

## V. CONCLUSION

In this study, a new method based on effective connectivity using Granger causality and directed transfer function was able to successfully extract connectivity patterns related to positive, negative and neutral emotions and led to the detection of these emotions based on EEG signal analysis. The obvious advantage of our method is its transparency in all stages of analysis and not using black boxes related to machine learning and neural networks, which increases its clinical applicability compared to previous works. However, this method needs further validation using different databases. In addition, our proposed method should be integrated with an artificial intelligence system to automate emotion recognition. In this

study, only three emotions were investigated, and future studies should evaluate this method for other emotions as well.

## ACKNOWLEDGMENT

This work was supported by the 2022 Guangxi University Young and Middle-aged Teachers' Basic Research Ability Improvement Project: Research and Practice of Applied Undergraduate Colleges Serving Local Economic Development - Taking Guangxi Normal University of Science and Technology as an Example (No.: 2022KY0847).

## REFERENCES

- [1] B. A. Erol, A. Majumdar, P. Benavidez, P. Rad, K.-K. R. Choo, and M. Jamshidi, "Toward artificial emotional intelligence for cooperative social human-machine interaction," *IEEE Transactions on Computational Social Systems*, vol. 7, no. 1, pp. 234-246, 2019.
- [2] L. F. Barrett, R. Adolphs, S. Marsella, A. M. Martinez, and S. D. Pollak, "Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements," *Psychological science in the public interest*, vol. 20, no. 1, pp. 1-68, 2019.
- [3] M. P. Herzberg and M. R. Gunnar, "Early life stress and brain function: Activity and connectivity associated with processing emotion and reward," *NeuroImage*, vol. 209, p. 116493, 2020.
- [4] A. Dzedzickis, A. Kaklauskas, and V. Bucinskas, "Human emotion recognition: Review of sensors and methods," *Sensors*, vol. 20, no. 3, p. 592, 2020.
- [5] X. Zhong, Y. Gu, Y. Luo, X. Zeng, and G. Liu, "Bi-hemisphere asymmetric attention network: recognizing emotion from EEG signals based on the transformer," *Applied Intelligence*, pp. 1-17, 2022.
- [6] M. Stanković and M. Nešić, "Functional brain asymmetry for emotions: Psychological stress-induced reversed hemispheric asymmetry in emotional face perception," *Experimental Brain Research*, vol. 238, no. 11, pp. 2641-2651, 2020.
- [7] A. Khaleghi, M. R. Mohammadi, K. Shahi, and A. M. Nasrabadi, "Computational neuroscience approach to psychiatry: A review on theory-driven approaches," *Clinical Psychopharmacology and Neuroscience*, vol. 20, no. 1, p. 26, 2022.
- [8] A. Afzali, A. Khaleghi, B. Hatf, R. Akbari Movahed, and G. Pirzad Jahromi, "Automated major depressive disorder diagnosis using a dual-input deep learning model and image generation from EEG signals," *Waves in Random and Complex Media*, pp. 1-16, 2023.
- [9] A. Khaleghi, M. R. Mohammadi, M. Moeini, H. Zarafshan, and M. Fadaei Fooladi, "Abnormalities of alpha activity in frontocentral region of the brain as a biomarker to diagnose adolescents with bipolar disorder," *Clinical EEG and neuroscience*, vol. 50, no. 5, pp. 311-318, 2019.
- [10] A. Khaleghi, M. R. Mohammadi, K. Shahi, and A. Motie Nasrabadi, "A neuronal population model based on cellular automata to simulate the electrical waves of the brain," *Waves in Random and Complex Media*, pp. 1-20, 2021.
- [11] A. Khaleghi, M. R. Mohammadi, K. Shahi, and A. M. Nasrabadi, "Possible Neuropathological Mechanisms Underlying the Increased Complexity of Brain Electrical Activity in Schizophrenia: A Computational Study," *Iranian Journal of Psychiatry*, pp. 1-7, 2023.
- [12] A. Khaleghi, A. Sheikhan, M. R. Mohammadi, and A. M. Nasrabadi, "Evaluation of cerebral cortex function in clients with bipolar mood disorder I (BMD I) compared with BMD II using QEEG analysis," *Iranian Journal of Psychiatry*, vol. 10, no. 2, p. 93, 2015.
- [13] A. Khaleghi et al., "EEG classification of adolescents with type I and type II of bipolar disorder," *Australasian physical & engineering sciences in medicine*, vol. 38, pp. 551-559, 2015.
- [14] M. Moeini, A. Khaleghi, N. Amiri, and Z. Niknam, "Quantitative electroencephalogram (QEEG) spectrum analysis of patients with schizoaffective disorder compared to normal subjects," *Iranian Journal of Psychiatry*, vol. 9, no. 4, p. 216, 2014.
- [15] M. Moeini, A. Khaleghi, and M. R. Mohammadi, "Characteristics of alpha band frequency in adolescents with bipolar II disorder: a resting-

- state QEEG study," Iranian journal of psychiatry, vol. 10, no. 1, p. 8, 2015.
- [16] M. Moeini, A. Khaleghi, M. R. Mohammadi, H. Zarafshan, R. L. Fazio, and H. Majidi, "Cortical alpha activity in schizoaffective patients," Iranian Journal of Psychiatry, vol. 12, no. 1, p. 1, 2017.
- [17] W. Xiao, G. Manyi, and A. Khaleghi, "Deficits in auditory and visual steady-state responses in adolescents with bipolar disorder," Journal of Psychiatric Research, vol. 151, pp. 368-376, 2022.
- [18] A. Khaleghi, P. M. Birgani, M. F. Fooladi, and M. R. Mohammadi, "Applicable features of electroencephalogram for ADHD diagnosis," Research on Biomedical Engineering, vol. 36, pp. 1-11, 2020.
- [19] A. Khaleghi, M. R. Mohammadi, G. P. Jahromi, and H. Zarafshan, "New ways to manage pandemics: Using technologies in the era of covid-19: A narrative review," Iranian journal of psychiatry, vol. 15, no. 3, p. 236, 2020.
- [20] A. Khaleghi, H. Zarafshan, and M. R. Mohammadi, "Visual and auditory steady-state responses in attention-deficit/hyperactivity disorder," European archives of psychiatry and clinical neuroscience, vol. 269, pp. 645-655, 2019.
- [21] M. R. Mohammadi, A. Khaleghi, A. M. Nasrabadi, S. Rafieivand, M. Begol, and H. Zarafshan, "EEG classification of ADHD and normal children using non-linear features and neural network," Biomedical Engineering Letters, vol. 6, pp. 66-73, 2016.
- [22] H. Zarafshan, A. Khaleghi, M. R. Mohammadi, M. Moeini, and N. Malmir, "Electroencephalogram complexity analysis in children with attention-deficit/hyperactivity disorder during a visual cognitive task," Journal of clinical and experimental neuropsychology, vol. 38, no. 3, pp. 361-369, 2016.
- [23] D. Dadebayev, W. W. Goh, and E. X. Tan, "EEG-based emotion recognition: Review of commercial EEG devices and machine learning techniques," Journal of King Saud University-Computer and Information Sciences, vol. 34, no. 7, pp. 4385-4401, 2022.
- [24] Y. Zhang, X. Ji, and S. Zhang, "An approach to EEG-based emotion recognition using combined feature extraction method," Neuroscience letters, vol. 633, pp. 152-157, 2016.
- [25] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," IEEE Transactions on autonomous mental development, vol. 7, no. 3, pp. 162-175, 2015.
- [26] M. Meng, Y. Zhang, Y. Ma, Y. Gao, and W. Kong, "EEG-based emotion recognition with cascaded convolutional recurrent neural networks," Pattern Analysis and Applications, vol. 26, no. 2, pp. 783-795, 2023.
- [27] S. Hwang, K. Hong, G. Son, and H. Byun, "Learning CNN features from DE features for EEG-based emotion recognition," Pattern Analysis and Applications, vol. 23, pp. 1323-1335, 2020.
- [28] R. Nawaz, K. H. Cheah, H. Nisar, and V. V. Yap, "Comparison of different feature extraction methods for EEG-based emotion recognition," Biocybernetics and Biomedical Engineering, vol. 40, no. 3, pp. 910-926, 2020.
- [29] N. S. Suhaimi, J. Mountstephens, and J. Teo, "EEG-based emotion recognition: A state-of-the-art review of current trends and opportunities," Computational intelligence and neuroscience, vol. 2020, 2020.
- [30] X. Li et al., "EEG based emotion recognition: A tutorial and review," ACM Computing Surveys, vol. 55, no. 4, pp. 1-57, 2022.
- [31] L. E. Ismail and W. Karwowski, "A graph theory-based modeling of functional brain connectivity based on eeg: A systematic review in the context of neuroergonomics," IEEE Access, vol. 8, pp. 155103-155135, 2020.
- [32] P. M. Rossini et al., "Methods for analysis of brain connectivity: An IFCN-sponsored review," Clinical Neurophysiology, vol. 130, no. 10, pp. 1833-1858, 2019.
- [33] M. C. Stevens, "The developmental cognitive neuroscience of functional connectivity," Brain and cognition, vol. 70, no. 1, pp. 1-12, 2009.
- [34] D. Goldenberg and A. Galván, "The use of functional and effective connectivity techniques to understand the developing brain," Developmental cognitive neuroscience, vol. 12, pp. 155-164, 2015.
- [35] K. J. Friston, "Functional and effective connectivity: a review," Brain connectivity, vol. 1, no. 1, pp. 13-36, 2011.
- [36] S. Bagherzadeh, K. Maghooli, A. Shalhaf, and A. Maghsoudi, "Emotion recognition using effective connectivity and pre-trained convolutional neural networks in EEG signals," Cognitive Neurodynamics, vol. 16, no. 5, pp. 1087-1106, 2022.
- [37] S. Bagherzadeh, K. Maghooli, A. Shalhaf, and A. Maghsoudi, "Recognition of emotional states using frequency effective connectivity maps through transfer learning approach from electroencephalogram signals," Biomedical Signal Processing and Control, vol. 75, p. 103544, 2022.
- [38] M. Abdolssalehi, A. Motie-Nasrabadi, M. Abdossalehi, and F. A. Motie Nasrabadi, "Combining independent component analysis with chaotic quantifiers for the recognition of positive, negative and neutral emotions using EEG signals," Indian Journal of Scientific Research, 2014.
- [39] T. Mullen, A. Delorme, C. Kothe, and S. Makeig, "An electrophysiological information flow toolbox for EEGLAB," Biol. Cybern, vol. 83, pp. 35-45, 2010.
- [40] M. Ding, S. L. Bressler, W. Yang, and H. Liang, "Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: data preprocessing, model validation, and variability assessment," Biological cybernetics, vol. 83, pp. 35-45, 2000.
- [41] L. Kilian and H. Lütkepohl, Structural vector autoregressive analysis. Cambridge University Press, 2017.
- [42] S. Koelstra et al., "Single trial classification of EEG and peripheral physiological signals for recognition of emotions induced by music videos," in Brain Informatics: International Conference, BI 2010, Toronto, ON, Canada, August 28-30, 2010. Proceedings, 2010: Springer, pp. 89-100.
- [43] Y.-P. Lin et al., "EEG-based emotion recognition in music listening," IEEE Transactions on Biomedical Engineering, vol. 57, no. 7, pp. 1798-1806, 2010.
- [44] G. Chanel, C. Rebetz, M. Bétrancourt, and T. Pun, "Emotion assessment from physiological signals for adaptation of game difficulty," IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, vol. 41, no. 6, pp. 1052-1063, 2011.
- [45] S. A. Hosseini and M. B. Naghibi-Sistani, "Emotion recognition method using entropy analysis of EEG signals," International Journal of Image, Graphics and Signal Processing, vol. 3, no. 5, p. 30, 2011.
- [46] S. Nasehi, H. Pourghassem, and I. Isfahan, "An optimal EEG-based emotion recognition algorithm using gabor," WSEAS transactions on signal processing, vol. 3, no. 8, pp. 87-99, 2012.
- [47] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," IEEE transactions on affective computing, vol. 3, no. 1, pp. 42-55, 2011.
- [48] M. Mikhail, K. El-Ayat, J. A. Coan, and J. J. Allen, "Using minimal number of electrodes for emotion detection using brain signals produced from a new elicitation technique," International Journal of Autonomous and Adaptive Communications Systems, vol. 6, no. 1, pp. 80-97, 2013.
- [49] G. Gainotti, "The role of the right hemisphere in emotional and behavioral disorders of patients with frontotemporal lobar degeneration: an updated review," Frontiers in aging neuroscience, vol. 11, p. 55, 2019.
- [50] E. D. Ross, "Differential hemispheric lateralization of emotions and related display behaviors: emotion-type hypothesis," Brain Sciences, vol. 11, no. 8, p. 1034, 2021.
- [51] E. H. Houssein, A. Hammad, and A. A. Ali, "Human emotion recognition from EEG-based brain-computer interface using machine learning: a comprehensive review," Neural Computing and Applications, vol. 34, no. 15, pp. 12527-12557, 2022.
- [52] W. Li, W. Huan, B. Hou, Y. Tian, Z. Zhang, and A. Song, "Can emotion be transferred?—A review on transfer learning for EEG-Based Emotion Recognition," IEEE Transactions on Cognitive and Developmental Systems, vol. 14, no. 3, pp. 833-846, 2021.
- [53] H. Liu, Y. Zhang, Y. Li, and X. Kong, "Review on emotion recognition based on electroencephalography," Frontiers in Computational Neuroscience, vol. 15, p. 84, 2021.
- [54] M. Maithri et al., "Automated emotion recognition: Current trends and future perspectives," Computer methods and programs in biomedicine, vol. 215, p. 106646, 2022.

# Development of an Image Encryption Algorithm using Latin Square Matrix and Logistics Map

Emmanuel Oluwatobi Asani<sup>1</sup>, Godsfavour Biety-Nwanju<sup>2</sup>, Abidemi Emmanuel Adeniyi<sup>3</sup>,  
Salil Bharany<sup>4</sup>, Ashraf Osman Ibrahim<sup>5</sup>, Anas W. Abulfaraj<sup>6</sup>, Wamda Nagmeldin<sup>7</sup>

Department of Computer Science, Landmark University, Nigeria<sup>1,2</sup>  
Landmark University SDG 11 Group, Landmark University, Nigeria<sup>1</sup>

Department of Computer Sciences, Precious Cornerstone University, Ibadan, Nigeria<sup>3</sup>

Department of Computer Science and Engineering, Lovely Professional University, Phagwara, Punjab 144402, India<sup>4</sup>

Creative Advanced Machine Intelligence Research Centre, Faculty of Computing and Informatics,  
Universiti Malaysia Sabah, Jalan UMS, 88400 Kota Kinabalu, Sabah, Malaysia<sup>5</sup>

Department of Information Systems, King Abdulaziz University, P.O. Box 344, Rabigh; 21911, Saudi Arabia<sup>6</sup>

Department of Information Systems, College of Computer Engineering and Sciences,  
Prince Sattam bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia<sup>7</sup>

**Abstract**—The goal of this study was to develop a robust image cryptographic scheme based on Latin Square Matrix and Logistics Map, capable of effectively securing sensitive data. Logistics mapping is a comparatively strong chaos system which enciphers with an unpredictability that significantly reduces the chance of deciphering. Additionally, the Latin square matrix stands out for its uniform histogram distribution, thereby bolstering its encryption's potency. The consequent integration of these algorithms in this study was therefore grounded in the scientific rationale of establishing a strong and resilient cypher technique. The study provides a new chaos-based method and extends the application of the probabilistic approach to the domain of symmetric key image encryption. Permutation and substitution approaches of image encryption were deployed to address the issue of images volume and differing sizes. The issue of misplaced pixel positions in the image was also adequately addressed, making it an effective method for image encryption. The hybrid technique was simulated on image data and evaluated to gauge its performance. Results showed that the algorithm was able to securely protect image data and the private information associated with them, while also making it very difficult for unauthorized users to decrypt the information. The average encryption time of 184( $\mu$ s) on seven (7) images showed that it could be deployed for real-time systems. The proposed method obtained an average entropy of 7.9398 with key space of  $1.17 \times 10^{77}$  and an average avalanche effect (%) of 49.9823 confirming the security and resilience of the developed method.

**Keywords**—Image encryption; algorithm; logistics map; Latin square matrix; chaos technology

## I. INTRODUCTION

As a result of the massive development of digital information technology, an ever-increasing volume of image data is being generated and distributed over the interconnected networks of computing infrastructure. These images typically contain confidential information such as trade secrets, military secrets, confidential medical reports and other types of secrets. It is therefore critical to protect sensitive data contained in these images from unauthorized access; many industries and fields rely on the secure transmission of such data [1]. As a

result, image transmission security has emerged as one of the most pressing topics in information sciences [1].

The method that is utilized to protect the confidentiality of images is known as cryptography. Cryptography is a method for safely transmitting data and communications by utilizing certain keys. This ensures that only the receiver to whom the information is addressed is aware of the true content of the message, hence preventing unauthorized access [2, 3]. It has a significant impact on the conversations that take place over mobile phones, as well as on e-commerce, the sending of emails, the transmission of financial information, and other areas of an individual's day-to-day life. The prefix "crypt" in the word "cryptography" refers to something that is "hidden" or "written" [4]. Information is encrypted by the use of mathematical presumptions and algorithms in the field of cryptography. These techniques are used to encode information before it is delivered, making it harder to decipher the information from its original form. Cryptography provides privacy by ensuring that transmitted data isn't known by external parties, it is reliable because it ensures there is no form of modification during storage and transfer of data from the sender to the receiver [4].

Encryption is a method of data security that entails encoding the data in such a way that it can only be deciphered by those who have been granted permission to do so. Examples of data that can be encrypted include multimedia files and sensitive papers [3]. Encryption and decryption are both possible outcomes of the usage of cryptography. Image encryption refers to the process of hiding an image using an encryption algorithm in such a way that its private information is protected and it is unavailable to attackers or unauthorized users [5]. Images are encrypted for a variety of purposes, including identifying the image's source, securing copyright information, preventing piracy, and preventing individuals who shouldn't have access to them from viewing them. Image encryption allows images to be shared via e-mail, the internet and other transmission media without worrying about these images being seen by unauthorized users. According to [6], all internet-connected citizens share more than 1.8 billion images

per day. This shows how much sensitive data can be lost just in a day and therefore shows the importance of encrypting these images. The development of good encryption algorithms has resulted from the necessity to meet the security needs of digital images [7, 8]. Therefore, in this study, a Latin square matrix and logistics mapping cryptographic scheme for image encryption was designed, simulated and implemented.

The study contributes significantly in advancing the frontiers of cryptography through the innovative integration of chaos-based and probabilistic encryption techniques, ensuring that the enciphering and deciphering phases are error-resistant. Chaos-based encryption techniques are known for their ability to provide high levels of security due to their inherent unpredictability, while the integration of probabilistic techniques ensures that the crypto processes are not deterministic but involves randomness, thus enhancing their resilience against various attacks. One other notable feature of the proposed method is its ability to handle images of varying sizes. This scalability is crucial in practical applications where images may have different dimensions. The developed technique also achieves semantic security, meaning that even with knowledge of parts of the plain images, it is computationally infeasible to recover the key or glean meaningful information from it. In summary, this study introduces a novel chaos-based image encryption method with a probabilistic approach, addressing the unique challenges posed by image data. Statistical, computational and differential attack evaluations conducted on the developed algorithm further highlight the effectiveness of the proposed method. The developed encryption method is secure, as it has a large key space, a high level of sensitivity to both cipher keys and plain images, and no known weaknesses.

The remaining sections of this study are highlighted as follows: Section II discusses the related work by previous researchers in the area of image encryption. Section III gives the details of the methodology used in this study. Section IV gives the obtained results. Section V concludes the study while Section VI provides the recommendation.

## II. RELATED WORK

Patel & Thanikaiselvan [9] presented a new image encryption algorithm that used Latin Square and Machine Learning techniques. The algorithm first generated a chaotic sequence using neural network-based pseudo-random number generator. This sequence was then used to create encryption key images, which were XORed with the input image to produce the encrypted image. The proposed algorithm was iterated a finite number of times to generate a cipher image population. A genetic algorithm was then used to optimize the population and find the least correlated cipher image. The model was resistant to communication channel attacks, such as noise addition, cropping, and JPEG compression.

Wang et al. [10] proposed a new image encryption algorithm that used Latin square matrices. The algorithm first generated a chaotic sequence using a Lorenz system. This sequence was then used to create a Latin square matrix, which was used to permute the pixels of the input image. The permuted was then diffused using a logistic map. The simulation results from the proposed image encryption

algorithm showed that it outperformed many existing image security algorithms. The algorithm was also resistant to communication channel attacks. A further investigation of this study is recommended, as it showed promises in protecting sensitive data in several applications.

Zhang et al. [11] introduced and implemented a Latin Square and random-shift based chaotic image encryption scheme. In some contexts, it was also referred to as the LSRS algorithm. The LSRS algorithm made use of a structure that consisted of pixel scrambling, replacement, and bit scrambling. The generation of Latin squares during the encryption process was correlated to the chaotic sequence, and this generation contributed to an increase in the system's overall level of security. In this line of study, the difficulty of decoding the method increased as a result of the fact that each encrypted image corresponds to a Latin square lookup table. The results of the simulation validated the LSRS algorithm's reliability as well as its efficiency.

Xu et al., [12], presented an algorithm that made use of self-orthogonal Latin Squares as its basis. This algorithm was also referred to as SOLS. In one cycle of encryption, the algorithm used the "permutation-substitution permutation" mode and the entire encryption procedure was carried out by a single SOLS. This research demonstrated that the substitution operation could be protected from differential attack by using permutation procedures at both the top and bottom levels during a single round of encryption. Therefore, at the bottom level, both substitution and permutation operations contributed to the diffusion effect. The results of the complexity analysis and simulation in the study demonstrates that the newly created algorithm was both secure and efficient, which demonstrates that it is suitable for use in real-time applications.

Zhang & Chen [13] developed a technique of encryption based on Henon chaotic maps. This encryption algorithm employed a two-phase encryption strategy. In the first step, the original image was fused with a key image, and the process was repeated in the second step. The authors asserted that their encryption approach was suitable and more resistant to brute force attacks than other similar methods since they employed both the key-image and the plain image back.

## III. METHODOLOGY

This study provides a new chaos-based method and extends the application of the probabilistic approach to the domain of symmetric key image encryption. It has been carefully developed to make certain that the cipher text is unpredictable. This method of image encryption made use of two different approaches, namely permutation and substitution, to encrypt square images. Permutation is referred to as the process of shifting around the locations of an image's pixels and substitution as a method for modifying the values of the pixels that are adjacent to one another. This approach addresses the issue of encrypting images with voluminous data because it is probabilistic and can accommodate images of any size. It generates a key stream using randomness, which is subsequently used in various phases to carry out operations of permutation and substitution thereby achieving semantic security. The processes of permutation and substitution was carried out without the need to wrongly misplace the pixel

positions in the image, making it an effective method for image encryption.

The encryption process which includes a fusion of the Latin Square Encryption (involving random key generation, generation of Latin Squares, Latin Square whitening, Latin Square permutation) and Logistic map Encryption. This process is represented algorithmically in Algorithm 1, while Fig. 1 depicts the framework of the proposed hybrid model.

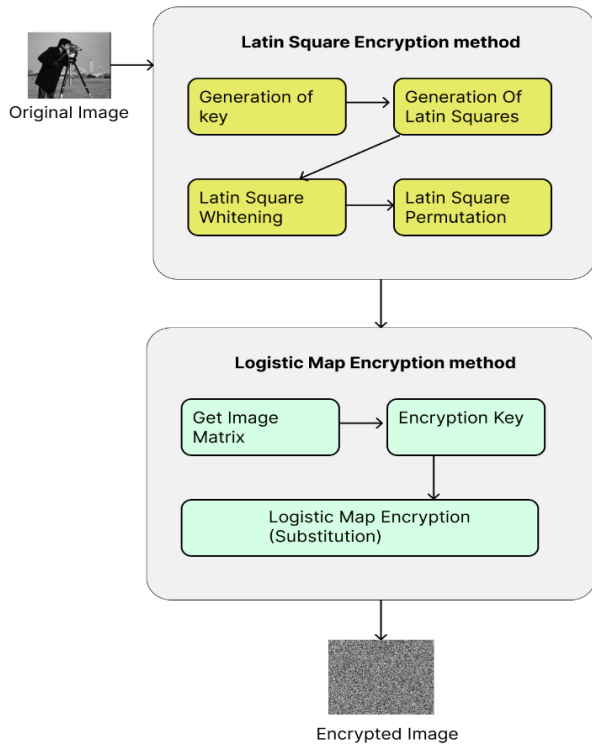


Fig. 1. Proposed fused latin square - logistic map encryption model.

**Algorithm 1: Fused Latin Square - Logistic map Encryption**

**Input:** p: Original image

**Output:** C: ciphered image

1. Import image and essential libraries;
2. Define **RandomKey()** to generate a 256-bit key;
3. Using predefined **KeyedLatin()** produce  $n$  Latin squares of order  $d$  dependent on the random key  $K$ ;
4. Define a **lsq\_whitening()** and use it to perform Latin square whitening;
5. Use the predefined **lsq\_permutation()** to randomly rearrange the image's pixel values;
6. Define **bitget(x,y)** to obtain the bits from each row and column and perform **XOR** operations on the obtained bits  $B$  and previous matrix  $M$ ;
7. Set **colourImage = 0** and **grayscaleImage = 1**;
8. Define **getImageMatrix(imageName)** for colour images and **getImageMatrix\_gray(imageName)** for grayscale images to obtain the image matrix.
9. Encrypted image for the first time with the **hexadecimal key sequence K**.
10. Conduct substitution on the image that has already been encrypted;
11. **Return Cipher Image C**

**A. Random Key Generation**

Algorithm 2 presents the method for generating random key using the *dec2hex* function.

**Algorithm 2: Random Key Generation Algorithm**

**Input:** image matrix

**Output:** k: a 256-bit random key in HEX

1. Define a function **dec2hex()**;
2. Parse **dec2hex()** using the string and length as parameters;
3. Convert decimal string to a hexadecimal string **hex\_string**;
4. Define function **RandomKey()**;
5. **Return key k.**

**B. Permutation Phase**

1) *Generating latin square*: A Latin square of order  $n$  is a square matrix with dimensions  $n * n$ , whose entries consist of  $n$  symbols ordered so that each symbol appears exactly once in each row and column. An  $n$  Latin square  $L$  of order  $d$  that is dependent on the key  $K$  is generated in this phase of the process. The mathematical modelling of a Latin square  $L$  of order  $d$  that can be derived using a tri-tuple function  $FL$  of  $(r, c, i)$  is presented as follows:

$$F_L(r, c, i) = \begin{cases} 1 & \text{si } L(r, c, i) = S_i \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Where  $r$  denotes the index of a line of an element of  $L$ ;  $l \in N = 0; 1; \dots; N - 1$ ,  $c$  denotes the index of a column of an element of  $L$   $c \in N$ ,  $i$  denotes the index of a symbol element in  $L$  and  $S_i$  is  $i$ th symbol in the whole =  $\{S_0, S_1, \dots, S_{N-1}\}$ . This implies that each symbol appears exactly once in each row and each column of  $L$ .

2) *Latin square whitening*: The term "Latin square Whitening" refers to an operation that combines the plaintext "P" with a pseudo random sequence. With the aid of the *LatinSquareWhitening* function, a  $n$  Latin Square  $L$  of order  $d$  that is determined by the key  $K$  is produced. An example is the *XOR* operation. When it comes to image encryption, a plaintext message is represented by an image block denoted by the letter  $P$  and made up of a certain number of pixels. Multiple binary bits (a byte) are used to represent each individual pixel. Since the goal of key whitening is to combine plaintext information with encryption keys, we describe it as an encrypted text that operates over a finite field  $GF(2^8)$  and is applied to the image data. A computational illustration of this is provided below:

$$y = [x + 1]_{2^8} \quad (2)$$

where  $x$  represents a byte in the plaintext,  $i$  is the byte's appropriate position in the keyed Latin square, and  $y$  represents the outcome of the whitening. The computations performed over  $GF(2^8)$  are denoted by  $[.]_{2^8}$ . The whitening effect caused by the procedure described above can be easily undone by using;

$$x = [y + 1]_{2^8} \quad (3)$$

Plaintext byte  $x$  represents a pixel in image encryption; for example, it might be identified at the intersection of the  $r$ th row and the  $c$ th column (i.e.,  $x = P(r, c)$ .)

Now let  $l = L(r, c)$  be an element situated at the relevant position in the keyed Latin square  $L$ , and let  $y$  be the ciphertext byte with  $y = C(r, c)$ , and then we will get the pixel-level equation below which is consistently utilized in the process of key whitening;

$$\begin{cases} C_{(r,c)} = [SR(P(r, c), [D]_3) + L(r, c)]2^8 \\ P_{(r,c)} = SR([C(r, c) + L(r, c)]2^8, [D]_3) \end{cases} \quad (4)$$

where the rotating parameter is  $D = L(0, 0)$ , symbol  $n$  indicates the existing round number ( $n[0,7]$ ), and  $SR$  represents the spatial rotating function ( $X, d$ ), which rotates an image  $X$  in accordance with various values of the direction  $d$  as defined below;

$$y = SR(X, d) \begin{cases} x, & \text{if } d = 0 \\ \text{flip } X \text{ up} \rightarrow \text{down} & \text{if } d = 1 \\ \text{flip } X \text{ left} \rightarrow \text{right} & \text{if } d = 2 \end{cases} \quad (5a)$$

Observe that the following identity is always true if  $Y = SR(X, d)$ :

$$X = SR(Y, d) \quad (5b)$$

3) *Latin square permutation*: At this phase, the Latin square is utilized to permute the image's pixels. If we take both the input and output  $x$  and  $y$  in FRM (Forward Row Mapping) and IRM (Inverse Row Mapping) to be indices, then a FRM defines a mapping from  $[0,1,2, \dots, 255]$  to  $[0,1,2, \dots, 255]$ , and an IRM defines the matching inverse mapping to that FRM mapping. Consequently, the Latin square P-box row, also known as the PLCL, can be defined as follows:

$$PLCL = \begin{cases} C_{(x,y_a)} = P(x, FRM(L, x, y_a)) \\ P_{(x,y_b)} = C(y, IRM(L, x, y_b)) \end{cases} \quad (6)$$

$y_a$  and  $y_b$  represent the column indices before and after the mapping, respectively, in this equation. In a comparable way, the P-column Latin square box, often known as the PCCL, can alternatively be calculated with the aid of the following equations:

$$PCCL = \begin{cases} C_{(x_a,y)} = P(x, FRM(L, x_a, y), y) \\ P_{(x_b,y)} = C(ICM(L, x_b, y), y) \end{cases} \quad (7)$$

In this operation, the row indices both before and after the mapping are shown by the notation  $x_a$  and  $x_b$ , respectively. Using a cascading method for the row permutations PLCLs, in addition to the column permutations known as PCCLs in the following manner allows us to construct our Latin square permutation with the highest level of performance possible:

$$\begin{cases} C_{(x,y)} = C^*(x, FCM(L, x, y), y) \\ P_{(x_a,y)} = P(x, FRM(L, x, y)) \end{cases} \quad (8)$$

In a broad sense, the function of the Latin square permutation can be expressed in the following manner:

$$PCL = \begin{cases} C = Ecr_p(L, P) \\ P = Dcr_p(L, P) \end{cases} \quad (9)$$

4) *Substitution using logistics maps*: The logistic map examines discrete time steps using a nonlinear difference equation. It is named the logistic map because it translates the value of the population at each given time step to its value at the subsequent time step.

The utilization of key mixing is included in this implementation. Following the completion of each pixel encryption, the initial values of the chaos map are recalculated with the addition of the key value as well as the value of the previous encryption. Logistics map is defined as follows;

$$\begin{aligned} x_{i+1} &= \alpha x_i (1 - x_i) + \beta y_i^2 x_i + \gamma z_i^3 \\ y_{i+1} &= \alpha y_i (1 - y_i) + \beta z_i^2 y_i + \gamma x_i^3 \\ z_{i+1} &= \alpha z_i (1 - z_i) + \beta x_i^2 z_i + \gamma y_i^3 \end{aligned} \quad (10)$$

$x_i, y_i$  and  $z_i$  represent system variables,  $\alpha$  and  $\gamma$  are parameters, while  $i$  shows iterations. For  $3.57 < \mu \leq 4$ , the map turns chaotic and for  $\mu = 4$ , the chaotic values produce in the full range of 0–1. The flowchart for decryption is shown in Fig. 2.

### C. Decryption

The decryption process converts the enciphered image back to plain image via the following algorithmic process.

- 1) Load the enciphered image;
- 2) Define *LogisticDecryption (imageName, key)*;
- 3) Generate Key-dependent 256x256 Latin Squares  $L$ ;
- 4) Extract a keyed Latin square;
- 5) Perform Latin square permutation using  $CP = lsq\_permutation(input, L, decryption)$ ;
- 6) Perform Latin square whitening utilizing  $CW = lsq\_whitening(input, L, decryption)$ ;
- 7) Return  $P$ ;

Fig. 2 presents the flowchart of the decryption process.

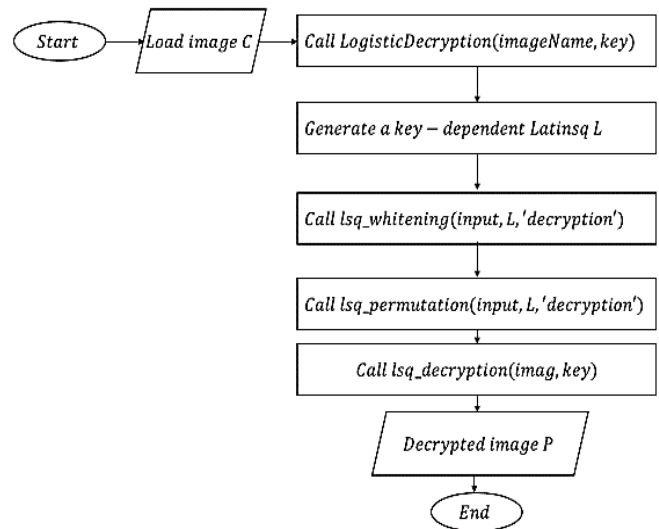


Fig. 2. Flow chart representation of the decryption process.

## IV. RESULTS AND DISCUSSIONS

### A. Simulation Results

The fused encryption algorithm was experimented on seven (7) image dataset (Baboon, Lena, Peppers, Man, Water Lilies, Airplane and Fruits) to demonstrate the performance of the technique in terms of complexity and security. The experimental results are presented as follows:

1) *Histogram analysis*: The act of making ciphertext more widely available is one that bears significant importance. To be more specific, it should conceal the excessive sections of the plain image and should not reveal any information about the image or the relationship between the image and the enciphered image. Additionally, it should conceal its redundant elements [14]. Algorithm 3 depicts the process of finding the image histogram.

---

#### Algorithm 3: Histogram Algorithm

---

**Input:** Image,

**Output:** Histogram graph.

1. Import cv2 and from matplotlib import pyplot as plt;
  2. set `img = cv2.imread('/image location path', 0)` to read the images from their location path;
  3. set `histr = cv2.calcHist([img], [0], None, [256], [0, 256])` to find the frequency of pixels in range 0 – 255;
  4. Display the plotting graph of an image using `plt.plot(histr)` and `plt.show()`;
- 

Fig. 3 and Fig. 4 contain the histograms of both the plain image and the encrypted form of those images respectively that were produced by the proposed approach. Both sets of histograms were generated by the proposed scheme. It is clear from the fact that the histograms of the cipher-images are relatively uniform, and they are notably different from that of the plain image. As a result, they do not present any signals that may be utilized to launch a statistical attack on the cipher.

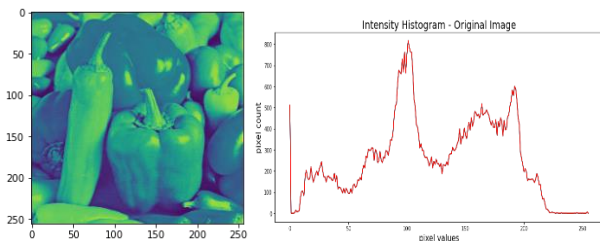


Fig. 3. Plain image and it's histogram.

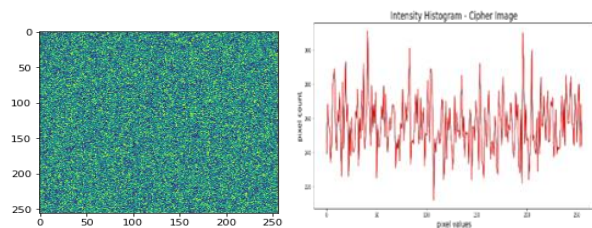


Fig. 4. Enciphered Fig. 3 and it's histogram.

2) *Key space analysis*: The larger the key space, the greater the ability to fight against an attack using brute force,

because the process of decryption is strenuous for the attacker, and the sophisticated nature of the information makes it impossible to retrieve [15]. However, if the key that is being encrypted is relatively simple, even the most robust encryption method can be broken using exhaustive search attack. This is not the case if the key is long [16]. The initial secret keys in the proposed approach were set to have a length of 256-bits. Because of this, the key space was 2256, which is equivalent to  $1.17 \times 10^{77}$ , making it sufficiently enormous to withstand any type of brute-force attack.

3) *An evaluation of the correlation between adjacent pixels*: A term referred to as an image's "intrinsic feature" describes the strong correlation that exists between the image's individual pixels. As a result, to increase resistance against statistical analysis, a secure encryption technique should remove it entirely. Within the scope of this work, visual autocorrelation analysis was conducted. Algorithm 4. depicts the process of plotting an auto-correlation pixel effect graph;

---

#### Algorithm 4: Auto-correlation pixel effect Algorithm

---

**Input:** Image,

**Output:** Histogram graph.

1. Import cv2, from matplotlib import pyplot as plt and pandas as pd;
  2. Declare data = pd.read\_csv("daily-minimum-temperatures-in-blr.csv",header=0, index\_col=0, parse\_dates=True, squeeze=True) to read the data from the csv;
  3. Display top 100 data using data.head(100);
  4. Display the plotting graph of an image using `pd.plotting.lag_plot(data, lag=1)`;
- 

Fig. 5 and Fig. 6 presents the autocorrelation results of both an original image and its enciphered version signposting the algorithms resilience to statistical analysis.

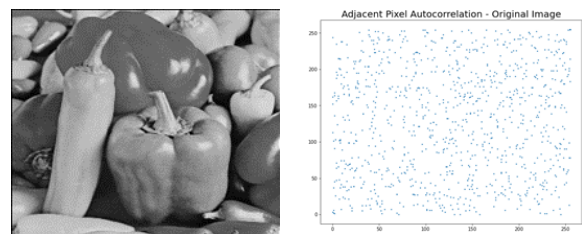


Fig. 5. Original image and its auto-correlation result.

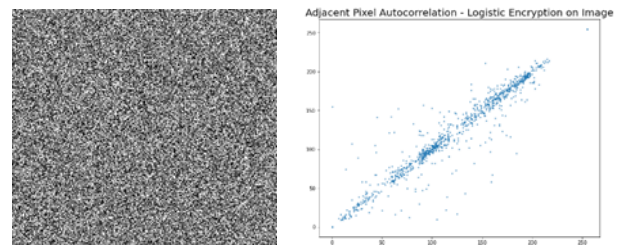


Fig. 6. Cipher image and its auto-correlation result.

4) *Execution time*: The execution time (ET) of the fused algorithm was computed programmatically using the following formula:

$$ET = \text{run time} + \text{compile time} \quad (11)$$

The average execution time result for this algorithm was calculated to be 184μs demonstrating the algorithms efficiency for real time applications. Table I presents the ET for seven images.

TABLE I. EXECUTION TIME OF ALGORITHM RESULTS

Image	ET(μs)
Baboon	233
Lena	210
Peppers	190
Man	120
Water Lilies	155
Airplane	146
Fruits	245
<b>Average ET</b>	<b>184</b>

5) *Mean Square Error (MSE)*: The MSE value is the average difference between the pixels throughout the entire image. It is utilized to determine how accurate the pixel value is, and the error value is just the difference between the two. A larger value for MSE denotes a greater degree of dissimilarity between the processed image and the original image. In spite of this, it is important to apply appropriate precautions when working with the edges. The mean square error (MSE) can be computed using the formula that is provided in the following equation.

$$MSE = \frac{1}{AB} \sum_{M=1}^{M=A} \sum_{N=1}^{N=B} (O(m, n) - R(m, n))^2 \quad (12)$$

In this case, AB refers to the size of the image, O refers to the image before it was processed, and R refers to the image after it was processed.

The MSE was calculated for the seven images in this study, and the calculation returns an average result of 0.0, which indicates that there was no error done on the edges of the images while they were being encrypted.

6) *Peak Signal Noise Ratio (PSNR)*: When comparing the squared error of the original image and the modified version, the Peak Signal to Noise Ratio (PSNR) and the Mean Square Error (MSE) are useful metrics to use. There is a connection that works reversely between PSNR and MSE. Therefore, a greater PSNR number suggests that the image has a higher quality (better). The ratio of the PSNR values of the decrypted and original versions is used to measure the image's quality. In this study, the PSNR value for the seven images experimented reached infinity, indicating a superior image quality. It has been determined, based on measurements, to be represented as:

$$PSNR = 10 \log_{10} \frac{(2^n - 1)^2}{MSE} \quad (13)$$

Algorithm 5 shows the step-by-step process of calculating PSNR

### Algorithm 5: PSNR Algorithm

**Input:** Original image, decrypted image,

**Output:** PSNR value.

1. **Import cv2 and math, from skimage import metrics;**
2. **Set  $img = cv2.imread('image location', 1)$   $dec\_img = cv2.imread('image location', 1)$  to read the images from their location path;**
3. **Assign  $img = dec\_img$ ;**
4. **Set  $psnr\_skimg = metrics.peak\_signal\_noise\_ratio(dec\_img, img, data\_rang=None)$**
5. **Print ('PSNR = ',  $psnr\_skimg$ )**

7) *Root Mean Square Error (RMSE)*: The RMSE value approximates the MSE value that provides accurate and reliable results [17]. Root mean square error (RMSE) is used to measure the difference between the original image and the segmented image [18]. It can be expressed in mathematical terms as;

$$RMSE = \sqrt{\frac{\sum_{l=1}^A \sum_{j=1}^B [Or(l,m) - De(l,m)]^2}{AB}} \quad (14)$$

where the values of the coordinates are denoted by the letters  $i$  and  $m$  and the size of the array is  $A \times B$ . Both the original and the decrypted versions of the image are indicated by the notation Or and De, respectively [19]. The interval  $[0, \infty]$  denotes the range of the RMSE. The smaller the value of Root Mean Square Error (RMSE), the more effective the segmentation. In this work, RMSE as calculated on the image datasets resulted to 0.0, which indicates that there is effective image segmentation while the images were being encrypted.

8) *Structural Similarity Index (SSIM)*: The SSIM demonstrates the correspondence between the decrypted and original image. This number is an evaluation and estimate of the image's quality that was produced from several different areas of the image that were the same size. SSIM is represented mathematically in Eq. (15).

$$SSIM = \frac{(2\mu_1\mu_{De} + C1)(2\partial_{1De} + C2)}{(\mu_1^2 + \mu_{De}^2 + C1)(\partial_1^2 + \partial_{De}^2 + C2)} \quad (15)$$

Here,  $\mu_1$  symbolizes the average of the inputs (I), while  $\mu_{De}$  represents the images after they have been decrypted (De). The standard deviations of the I and the De are, respectively,  $\partial_1^2$  and  $\partial_{De}^2$ . " $\partial_{1De}$ " stands for the covariance of the values I and De, while " $C1$ " and " $C2$ " stand for the regularization using the values  $(0.01P)^2$  and  $(0.01P)^2$ , respectively. The results of the SSIM can range from 0 to 1, with 1 indicating an excellent match between the original image and the image that has been modified. Moreover, good encryption algorithms should have SSIM values that fall anywhere between 0.97 and 1. In this study, the average SSIM value is 1 indicating that there is an exact match between the two images.

9) *Relative Average Spectral Error (RASE)*: RASE is a method that is used to calculate the overall performance of



image fusion algorithms for each spectral band. The following is the formula that is used to calculate RASE:

$$RASE = \frac{100}{A} \sqrt{\frac{1}{L} \sum_{i=1}^L RMSE^2 (B_i)} \quad (16)$$

x is the total exposure value of the L band (Bi) in the original multispectral image, whereas RMSE is the minimal square error that evaluates the effectiveness of each band in the merged image. The perfect value would be 0. In this study, RASE was calculated to be 0, which indicates that the methodology used was effective.

10)Relative dimensionless global error in synthesis (ERGAS): ERGAS is a metric that determines the overall quality of the merged image. The inaccuracy when the quality of anything improves, shows that there is a substantial tendency for ERGAS to decrease. As a result, it is an effective measure of the quality. Cases that are considered to be of "high quality" have error ERGAS values that are equal to or lower than 3, whereas cases that are considered to be of "poor quality" have error ERGAS values that are greater than 3. 0 is the optimum value for it. The average ERGAS of this work was computed to be 0.057143. Table II shows results of the analysis. This demonstrates that the algorithm is efficient. It is denoted mathematically as:

$$ERGAS = 100 \frac{he}{l} \sqrt{\frac{1}{L} \sum_{i=1}^n \frac{RMSE^2(A_i)}{y_i}} \quad (17)$$

TABLE II. ERGAS RESULTS

Image	ERGAS
Baboon	0.00
Lena	0.10
Peppers	0.20
Man	0.50
Water Lilies	1.00
Airplane	0.00
Fruits	0.00
Avg ERGAS	0.057143

11)Information entropy analysis: The entropy of the information is a measure of how random the information is, and it may be computed as follows [19]:

$$H(m) = \sum_{i=0}^{255} -Pl(m_i) \log_2 Pl(m_j) \quad (18)$$

Each gray level in an image with 256 gray levels contains 8 bits of data associated with it [20]. When the probability of each gray level is the same, the encrypted image is able to obtain the ideal entropy of 8, which indicates that each gray level of the encrypted image is evenly distributed. The fused encryption's entropy analysis on the enciphered images showed that the average information entropy was 7.53302, which is relatively close to the number 8. Hence, cipher images have a stronger random distribution, and the risk of information disclosure is completely eliminated. Table III depicts the results of evaluation.

TABLE III. AVERAGE OF INFORMATION ENTROPY RESULTS

Image	Entropy
Baboon	7.8693
Lena	7.9976
Peppers	7.9943
Man	7.8997
Water Lilies	7.9907
Airplane	7.993
Fruits	7.8330
Avg entropy	7.9397

12)Differential attack analysis: In most cases, attackers will begin by making a few alterations to the plain image. Next, they will encrypt both the plain image and the modified plain image using the same encryption algorithm. Finally, they will compare the two cipher images in order to gain further insight into the connection between the plain image and the cipher image. The number of pixels change rate (NPCR) and the unified average change intensity (UACI) are calculated to enhance the anti-differential attack performance of the encryption algorithm. In order to test the effect of slightly changing the plain image on the corresponding cipher image, the equations for NPCR and UACI are as follows:

$$NPCR = \frac{\sum_{i=1}^A \sum_{j=1}^B D(i,j)}{A \times B} \times 100\% \quad (19)$$

$$UACI = \frac{\sum_{i=1}^A \sum_{j=1}^B |C_1(i,j) - C_2(i,j)|}{A \times B} \times 100 \quad (20)$$

If the height and breadth of the cipher image are denoted by A and B and C1, C2 are two cipher images, and there is a difference of one bit between each answering plain image and each cipher image. The values of NPCR and UACI that should be anticipated for an image are the following: NPCR = 99.6094% and UACI = 33.4094%, respectively.

During the simulation, we selected a pixel at random for each of the image data in order to alter the value of that pixel. After that, we encrypted the images both before and after the change in order to obtain two cipher images, then we computed the NPCR and UACI. It was deduced, based on the simulation result, that the NPCR of the encrypted image was 99.60983 % and that the UACI was 33.4235 %, both of which are extremely similar to the value that was anticipated. Hence, the encryption technique is sensitive to alterations and has the ability to withstand differential attack. Table IV indicates result of the analysis.

TABLE IV. RESULTS OF NPCR AND UACI ANALYSIS

Image	NPCR	UACI
Baboon	99.6087	33.5866
Lena	99.6047	33.4082
Peppers	99.6076	33.2532
Man	99.6087	33.8456
Water Lilies	99.6067	33.2659
Airplane	99.6077	33.2418
Fruits	99.6083	33.4235
Avg entropy	99.6074	33.43211

13) *Avalanche effect*: To examine key sensitivity and demonstrate the robustness of the suggested scheme against differential cryptanalysis, evaluations on avalanche properties were conducted. In [20], it was stated that the avalanche effect of the technique should always be  $\geq 50\%$ . As part of this study, the avalanche effect of encryption algorithm was evaluated. The pattern of the number of bits modified in cipher with a single bit change in the secret key revealed that, regardless of the position of the key bit altered, the average change in the number of bits in the cipher text was 49.98235%

14) *which is roughly 50%*. It shows in Table V that the proposed method has high key sensitivity, high confusion, and consistent and significant contributions from all key bits to the cipher bits. Algorithm 6 depicts the step-by-step process of this security analysis. Avalanche effect is calculated with the formula below:

$$\text{Avalanche Effect} = \frac{\text{No. of altered bits in the ciphertext}}{\text{No. of bits in the ciphertext}} \quad (21)$$

**Algorithm 6: Avalanche Effect Algorithm**

**Input:** Original image,

**Output:** Avalanche value.

1. **Def**  $x_0$  as first cipher;
2. **Def**  $x_1$  as second cipher after 1 bit change;
3. **Print bitwise xor operation;**
4. **Count 1s in binary;**
5. **Evaluate** equation of avalanche effect;
6. **Divide** result in step 5 by the longest **binary string**.

Table VI presents a comparison of this study with other state-of-the-art techniques, based on multiple evaluation metrics, while Fig. VII presents a comparative analysis of our techniques' encryption time with other authors. The comparison showed that our technique efficient and competes very favourably against other state-of-the-art techniques.

The plot displays the encryption time of the proposed method at 184( $\mu$ s) and plotted against various authors in the related work. The plot shows that the developed model achieves second best encryption time when compared to other state-of-the-art results. Fig. 8 shows the plot of average avalanche effect obtained from the experiment.

TABLE V. AVALANCHE EFFECT RESULTS FOR EACH BIT CHANGED IN THE SECRET KEY

Image	paswd (%)	pstd (%)	qwd (%)
Baboon	50.0208	49.9652	49.9350
Lena	50.0101	49.9916	49.9764
Peppers	50.0258	50.0204	49.9737
Man	49.7809	50.0052	49.9801
fruits	50.0600	50.0015	49.9448
Water lilies	49.9740	50.0458	49.9770
Airplane	49.9862	49.9702	49.9845
Average	49.97969	49.99999	49.9673

Average Avalanche effect = 49.98235%

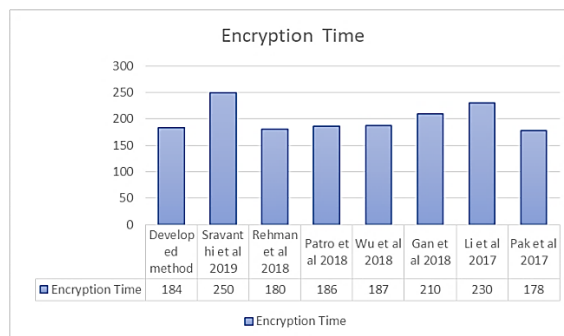


Fig. 7. Plot of encryption time and various authors result.

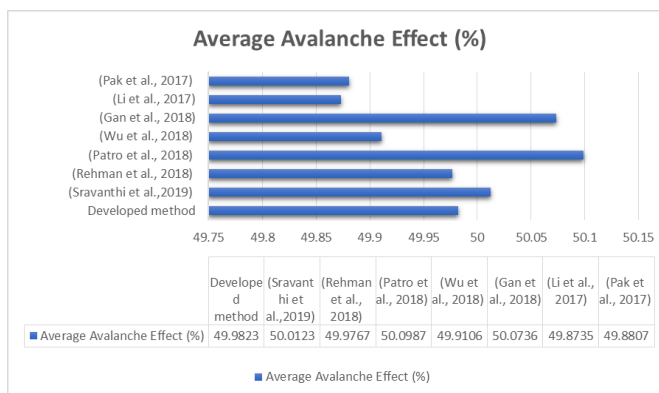


Fig. 8. Plot of average avalanche effect obtained from the experiment.

TABLE VI. COMPARISON OF SIMULATION RESULTS WITH RELATED SCHEMES

Author(s)/Reference	Average Entropy	Key Space	Average NPCR (%)	Average UACI (%)	Average MSE	Average RMSE	Average SSIM	Execution Time ( $\mu$ s)	Average Avalanche Effect (%)
Developed method	7.9398	$1.17 \times 10^{77}$	99.6074	33.4321	1.00	0.0	1.0	184	49.9823
(Sravanthi et al.,2019)	7.9993	$1.1 \times 2^{377}$	99.6098	33.4707	0.97	0.0	0.3	250	50.0123
(Rehman et al., 2018)	7.6635	$10^{94}$	99.5999	33.3848	1.00	0.0	0.0	180	49.9767
(Patro et al., 2018)	7.9998	$1.9 \times 2^{426}$	99.6028	33.4021	0.99	0.0	0.1	186	50.0987
(Wu et al., 2018)	7.9196	$10^{88}$	99.6090	33.4227	1.00	0.0	0.2	187	49.9106
(Gan et al., 2018)	7.9993	2470	99.6000	33.4400	0.96	0.0	0.1	210	50.0736
(Li et al., 2017)	7.9272	2273	99.6100	33.4600	0.95	0.0	1.0	230	49.8735
(Pak et al., 2017)	-	2138	99.6236	33.3441	0.99	0.0	1.0	178	49.8807

The avalanche effect is considered as one the desirable property of any encryption algorithm. The effect ensures that an attacker cannot easily predict a plain-text through a statistical analysis.

## V. CONCLUSION

The study's objective was to develop a resilient encryption scheme to protect images from being decrypted. A strong algorithm for encrypting images should be resistant to attacks, and its efficiency should be unaffected by either the encryption key or the image that has been encrypted. It should have a large key space. This algorithm with all its component parts and stages, satisfies each one of these criteria, and it also has certain unique qualities. As this proposed algorithm is completely described in integers, it is computationally efficient in either software or hardware and does not lead to complications with finite precision or discretization. In addition, the suggested approach builds all encryption primitives based on a key generator. These encryption primitives include substitution and permutation, and because of changes, the proposed method achieves high sensitivity to any key change. The suggested technique additionally combines probabilistic encryption, which provides the conversion of a single plain image into numerous cipher images utilizing a single encryption key, and assures that the decoding phase is error-resistant up to a preset level. Statistical, computational and differential attack evaluations have been conducted on the developed algorithm. All experimental results indicated that the proposed encryption method is secure, as it has a large key space, a high level of sensitivity to both cipher keys and plain images, and no known weaknesses.

## VI. RECOMMENDATION

It is strongly suggested that this method be utilized whenever images need to be encrypted. In addition, more research can be done on the algorithm to enhance its current capabilities and make it even more powerful than it now is. Additional statistical analysis can be carried out computationally as well as visually.

## REFERENCES

- [1] Y. L. Hailan Pan, "Research On Digital Image Encryption Algorithm Based On Double Logistic Chaotic Map". Research on digital image encryption algorithm based on double logistic chaotic map, 10. 2018.
- [2] M. Xu, and Z. Tian, "A Novel Image Encryption Algorithm Based on Self-Orthogonal Latin Squares". Optik, vol. 17 no. 1, pp. 891-903, 2018.
- [3] A. E. Adeniyi, S. Misra, E. Daniel, and A. Bokolo Jr, "Computational complexity of modified blowfish cryptographic algorithm on video data". Algorithms, vol. 15 no. 10, pp. 373. 2022.

- [4] L. Abraham, and N. Daniel, "Secure Image Encryption Algorithms: A Review". INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH, vol. 2. 2013
- [5] A. E. Adeniyi, K. M. Abiodun, J. B. Awotunde, M. Olagunju, O. S. Ojo, and N. P. Edet, "Implementation of a block cipher algorithm for medical information security on cloud environment: using modified advanced encryption standard approach". Multimedia Tools and Applications, pp. 1-15. 2023.
- [6] M. Meeker's, "KPCB Researcher Mary Meeker's Annual Internet Trends Study". KPCB researcher Mary Meeker's annual Internet Trends study. 2017.
- [7] Q. R. Renzhi Li, "Novel image encryption algorithm based on improved logistic map". Novel image encryption algorithm based on improved logistic map, vol. 10, 2019.
- [8] A. E. Adeniyi, A. L. Imoize, J. B. Awotunde, J. B., C. Lee, P. Falola, R. G. Jimoh, and S. A. Ajagbe, "Performance Analysis of Two Famous Cryptographic Algorithms on Mixed Data". Journal of Computer Science, vol.19, no. 6, pp. 694-706. <https://doi.org/10.3844/jcssp.2023.694.70> 2023.
- [9] S. Patel, and V. Thanikaiselvan, "Latin Square and Machine Learning Techniques Combined Algorithm for Image Encryption". Circuits, Systems, and Signal Processing, pp. 1-25, 2023.
- [10] X. Wang, Y. Su, M. Xu, H. Zhang, and Y. Zhang, "A new image encryption algorithm based on Latin square matrix". Nonlinear Dynamics, vol. 10 no. 7, pp. 1277-1293. 2022.
- [11] X. Zhang, T. Wu, Y. Wang, L. Jiang, and Y. Niu, "A novel chaotic image encryption algorithm based on latin square and random shift". Computational Intelligence and Neuroscience, 2021.
- [12] M. Xu, & Z. Tian, "A Novel Image Encryption Algorithm Based on Self-Orthogonal Latin Squares". Optik, vol. 17 no. 1, pp. 891-903. 2018.
- [13] X. Zhang, and W. Chen, "A new chaotic algorithm for image encryption". In 2008 International conference on audio, language and image processing, IEEE, pp. 889-892. 2008.
- [14] M. Jiang, P. Cui, and C. Faloutsos, "Suspicious behavior detection: Current trends and future directions". IEEE intelligent systems, vol. 31, no. 1, pp. 31-39, 2016.
- [15] A. E. Omolara, A. Jantan, O. I. Abiodun, K. V. Dada, H. Arshad, and E. Emmanuel. "A deception model robust to eavesdropping over communication for social network systems". IEEE Access, vol. 7, pp. 100881-100898. 2019.
- [16] F. Han, X. Liao, B. Yang, and Y. Zhang, "A hybrid scheme for self-adaptive double color-image encryption". Multimedia Tools and Applications, vol. 77, pp. 14285-14304. 2018.
- [17] B. M. Lei Chen "Differential cryptanalysis of a novel image encryption algorithm based on chaos and Line map". Nonlinear Dynamics, 12, 2016.
- [18] M. K. Kumar. "Colour Image Encryption Technique Using Differential Evolution In Non-Subsampled Contourlet Transform Domain". IET Image Processing, 11. 2018.
- [19] M. Kaur, and V. Kumar "A Comprehensive Review on Image Encryption Techniques". Archives of Computational Methods in Engineering, vol. 27, no.1, pp. 15-43, 2020.
- [20] X. W. Hao Zhang, "An Efficient and Secure Image Encryption Algorithm Based on Non- Adjacent Coupled Maps". 17. 2020

# Tampering Detection and Segmentation Model for Multimedia Forensic

Manjunatha S<sup>1</sup>, Malini M Patil<sup>2</sup>, Swetha M D<sup>3</sup>, Prabhu Vijay S S<sup>4</sup>

Dept. of Information Science & Engineering, Global Academy of Technology (Affiliated to Visvesvaraya Technological University, Belagavi-590018), Bengaluru, Karnataka, India<sup>1</sup>

Dept. of Computer Science & Engineering, RVITM (Affiliated to Visvesvaraya Technological University, Belagavi-590018), Bengaluru, Karnataka, India<sup>2</sup>

Dept. of Computer Science & Engineering, BNMIT (Affiliated to Visvesvaraya Technological University, Belagavi-590018), Bengaluru, Karnataka, India<sup>3</sup>

Dept. of Information Science & Engineering, BMSCE, Bengaluru, Karnataka, India<sup>3</sup>

Senior Software Engineer and Data Analyst, Navshyatechnologies, Bengaluru, Karnataka, India<sup>4</sup>

**Abstract**—When an image undergoes hybrid post-processing transformation, detecting tamper region, localizing it and segmentation becomes very difficult tasks. In particular, when a copy-move attack with hybrid transformation has similar contrast and illumination parameters with an authenticated image it makes tamper detection difficult. Alongside, under small-smooth attack existing tamper identification model provides a very poor segmentation outcome and sometimes fails to identify an image as tampered. This article focused on addressing the difficulty through the adoption of the Deep Learning model. The proposed technique is efficient in detecting tampering with good segmentation outcomes. However, existing models fail to distinguish adjacent pixels' relationships affecting segmentation outcomes. In this paper, an Improved Convolution Neural Network (ICNN) assuring correlation awareness-based Tamper Detection and Segmentation (TDS) model for image forensics is presented. This model brings good correlation among adjacent pixels through the introduction of an additional layer namely the correlation layer alongside vertical and horizontal layers. The TDS-ICNN is very effective in localizing and segmenting tamper regions even under small-smooth post-processing tampering attacks by using a feature descriptor built using aggregated three-layer ICNN architecture. An experiment is done to study TDS-ICNN with other tamper identification models using various datasets such as MICC, Coverage, and CoMoFoD. The TDS-ICNN is very efficient under different post-processing hybrid attacks when compared with existing models.

**Keywords**—Convolution neural networks; digital image forensic; hybrid image transformation; resampling feature; segmentation

## I. INTRODUCTION

Image authentication methods are characterized in the following two classes: (1) Active and (2) Blind or Passive. Digital Watermarking has been proposed as an active method using which an image can be authenticated [1]. The main aim of watermarking is to ensure the protection of copyright, authentication of content, ownership recognition, and data integrity. Watermarking ensures content from modification only and also provides data integrity and content authentication. Watermarks generally are indivisible from the digitized picture element they are embedded in. Further, the

watermarks undergo a similar transformation in the picture. The major drawback of using watermarking is that it prerequisites watermark to be embedded during capturing of the image. This also binds/restricts its applicability to real-time environment usage. Thus, they are used only in controlled surroundings such as in armed forces and surveillance environments. Furthermore, some watermarks may break down the image quality.

Passive or blind forgery detection considers images without any digital signature, digital watermark, or any other prior information and checks the authenticity and origin of the image. Image forgery may not leave any visual clues of tampering being done. But there are high chance that it most probably perturbs the underlying statistical characteristics of an original image or modifies the scene of an image. These inconsistencies are utilized for tampering identification. Since this method doesn't require any prior knowledge of the picture. Passive forgery authentication techniques are further divided into forgery-dependent techniques and forgery-independent techniques as shown in Fig. 1. Forgery-dependent methodologies are delineated to identify a particular class of tampering for example splicing, copy-clone, etc. which relies on the forgery class type used on a picture. Whereas latter, the independent methods identify tampering using artifact traces left in the procedure of carrying out light inconsistencies and re-sampling. Existing forgery detection techniques recognize various traces of forged segments and identify them and the forged segment is localized [2].

Over many years, several attempts have been made for the classification of whether given images are authenticated or forged. Nonetheless, just a couple of works [3] endeavor to localize tampering at the pixel level. Recent methodologies [4] have aimed at addressing the localization issue by characterizing patches as tampered. Establishing the location of the tampered region ring is an exceptionally difficult job and also well-crafted tampering of pictures doesn't leave any visual hints. A sample repetition of well-crafted image tampering is shown in Fig. 1, where image one and three defines the tampered image, and image two and four defines the ground truth of the respective image that has been forged through transformation attacks. In Fig. 1(a), a copy-clone

attack is presented where a set of objects is copied and pasted into the different regions within the image. Here one image is the source and the other is the copy-moved object. Fig. 1(b), defines the spliced attack, here an object within an image is spliced and copied into a different image. Fig. 1(c), shows an object removal attack, where an image is blended on top of some-other object.

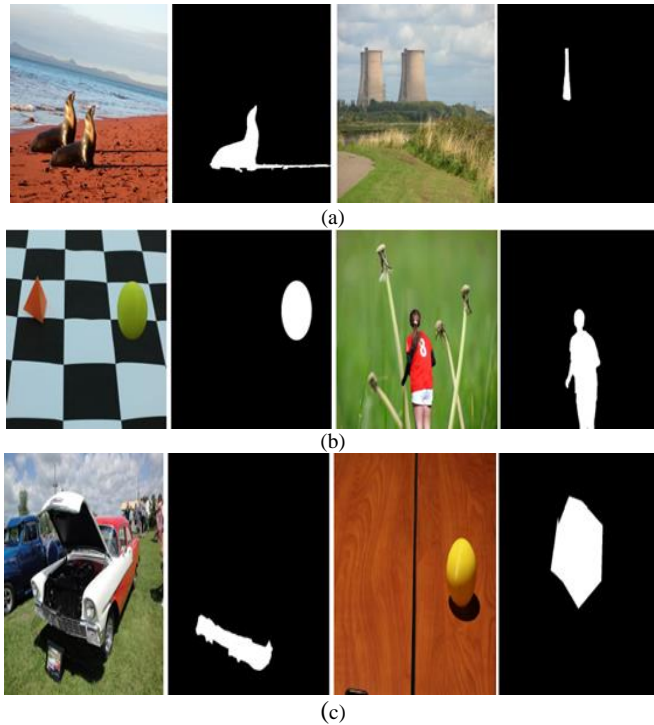


Fig. 1. Different types of tampering attacks [7].

The majority of the cutting-edge image forgery detection approaches uses statistical properties through frequency domain feature. In [5], the artifact was introduced by applying a different level of JPEG compression for the identification of tampered images. In [6], additional noisy information was added to images that were compressed through JPEG to work on the presentation of resampling identification. Recently, the DL method has provided some very good results in computer vision applications, for example, object detection, hyperspectral crop classification, image registration, and segmentation, etc. Recently, DL models like auto-encoders [7] and Convolutional Neural Networks (CNN) [8], [9] have been employed for image tampering detection with good results. Existing tampering detection models are predominantly designed to detect only one kind of attack [10], [11]. In this way, one methodology probably won't excel in different sorts of tampering attacks. Additionally, it appears to be not realistic to expect this sort of attack well in advance.

Segmenting the tampering region is more challenging as compared to object segmentation because here only the region that is tampered only must be segmented. Recently, CNN has been emphasized with good effect for object segmentation strategies [12], [13]. In [12], a fully connected CNN has been used for studying object features and shape features through the extraction of features in a hierarchical manner. The CNN-based method provides good performance in the field of

segmentation and object classification. In image tampering only the tampered region must be segmented and well-crafted tampered image differentiating between genuine and tampered is very difficult because they look very comparable. Although CNN produces spatial guides for various districts of sections, it can't sum up some different statistical noise made by various tampering methods. Consequently, the tampering region localization using a standard CNN-based design may not provide the ideal performance requirement of a realistic attack. In [13], studied different image forgery segmentation models were studied [14]. The study shows that the existing model performs badly in detecting copy-clone and object-removal. Using the resampling feature [4] the artifacts were created (i.e., resampling, compression) using tampered images can be learned [15]. The resampling attack generally occasionally allows correlation among pixels because of interpolation. The CNN-based [16], [17], image forgery identification model learns resampling features [18] very well using spatial maps produced through translation invariance of various regions of images [19], [20]. Thus, this research work aims to build an efficient resampling feature detection through CNN to detect hybrid attacks and achieve better-tampered region segmentation outcomes [21], [22]. The significance of the research work is as follows:

- This paper presented an improved CNN for tampering detection and segmentation in the image by adding to additional layer to retain the correlation between horizontal and vertical streams for exploiting good-quality resampling features.
- The TDS-ICNN model can work well considering different attacks such as scaling, compression, and rotation attacks.
- The TDS-ICNN is efficient in detecting multiple tampered regions within the same image.
- The TDS-ICNN can even detect image tampering attacks under noisy and small-smooth regions. An improved tampering area segmentation outcome using TDS-ICNN for tampering dataset with hybrid transformation attack is achieved. On the other side, the existing model works well i.e., good segmentation for some datasets, and for other datasets, very poor result is achieved.
- This shows the robustness of the TDS-ICNN model. An improved ROC performance is achieved using the TDS-ICNN model for carrying out classification tasks such as whether a given image is authenticated or tampered with considering diverse tampering datasets such as CoMoFoD, Coverage, and MICC.

The manuscript is arranged as follows: Section II discusses various current methodologies to detect tampering in multimedia content. Section III presents the material and method used for performing tampering detection methods. Section IV presents with working structure of the proposed tampering detection and segmentation model. Section V presents the experiment analysis of the proposed method with various other tampering detection methodologies. Section VI

concludes the research significance with future research direction.

## II. RELATED WORK

The section studies various recent methodologies for detecting tampering in multimedia content. In [8] developed a robust image tampering detection method using CNN, where an image undergoes double compression tampering attacks; the model attains an accuracy of 92% using the CASIA v2 dataset. Similarly, [9] used ResNet50v2 for constructing batchwise CNN to detect image tampering. Experiment outcomes show 99.3 accuracy on the Casia v1 dataset and 81% accuracy on the CASIA v2 dataset. In [11] designed a tampering detection by training CNN with both unseen noise and predictable noise for online social network platforms. The model works well for social platforms; however, considering other domains the model fails to accurately detect tampering in images. In [18] designed pulse-CNN model to extract the contour features of potential tampering that had undergone complex tampering attacks like noise, scaling, and rotation attacks. The experiment outcome shows the model achieved a precision of 95.27% and 95.3% on the CASIA and CoMoFoD datasets, respectively.

In [23] introduced an end-to-end deep neural network namely BusterNet with two layers to capture the tamper feature followed by a fusion layer to merge the feature for segmentation of copy-move tamper region. Experiments are done on CASIA and CoMoFoD and segmentation output is given at pixel-level. Similarly, in [26] designed adaptive-attention residual refined network (AR-Net) to extract tampered object features, and feature maps correlation is done after which the fusion of features is performed using pyramid pooling. The experiment is done using CASIA II, Coverage, and CoMoFoD. In [27] developed a copy-move tampering detection mechanism using source-target region distinguishment network (STRDNet) by extending BusterNet. The model additionally introduces a filter at the pooling layer with a double self-correlation layer for establishing feature matching hierarchically. The experiment is done using CASIA, CoMoFoD, and Coverage datasets and the segmentation outcome is given at the pixel level. In [29] introduced an effective block-level feature optimization trained with deep CNN. The deep CNN uses a feature pyramid for robust detection accuracy against scaling attacks. The experiment is done using CASIA II with 57.48% and the CoMoFoD dataset with a precision of 50.11% and the boundary pixel direction aids in the detection of segmentation edges and can tolerate noise, compression, blurring, and color addition.

In [24] designed a key-point-based clustering method to detect tampering attacks under small-smooth regions. Experiments are done on MICC, GRIP, FAU, and Coverage with good true positive rates of 97.5%, 100.0%, 100%, and 80.22%, respectively. In [25] designed a new SIFT key points extraction through effective clustering for identifying

tampered regions utilizing similarities. The clustering process to identify similarities is done considering color with different scales and smaller cluster size is considered to reduce computational overhead. In obtaining more quality outcomes pixel level similarity is done iteratively. The experiment is done using D0 datasets and pixel-level analysis segmentation accuracy is measured. In [30] combined both accelerated KAZE (A-KAZE) and speeded up robust features (SURF) for extraction of features by keeping the contrast level reasonably low. Then, to eliminate the mismatch density-based spatial clustering (DBSCAN) is used. Then, the affine matrix is applied to improve the tampering localization accuracy. The experiment is done with Ardizzone (D0) with 92.75% precision and the CoMoFoD dataset [31] with 95.23%. The overall survey shows key-point-based tampering detection is predominantly studied its performance using the MICC and DO dataset and the CNN-based model is predominantly studied using CASIA, Coverage, and CoMoFoD dataset.

The result attained using existing tampering detection methods have obtained satisfactory results; however, there is still wide scope to improve the results, as the existing model failed to provide good segmentation result under small-smooth robust tampering attacks which undergoes diverse post-processing attacks. Further, the model must be tested under different kinds of datasets; and most of the existing methods failed to provide pixel-level segmentation analysis. The current methods failed to extract feature correlation between horizontal and vertical layers; as a result, higher false positive is experienced with poor segmentation outcomes. In overcoming the research issues in the next section, the proposed methodology is presented.

## III. MATERIALS AND METHODS USED

### A. Dataset Used

The dataset used in this work is listed below:

1) *MICC*: The dataset is composed of 600 images out of which, 160 images are forged, and the remaining 440 images are authenticated. The dataset is composed of different attacks like scaling and rotation made of plants, artifacts, animals, etc.

2) *CoMoFoD*: The dataset is made of a total of 260 images where several post—processing attacks are done. The resolution images are 512×512 for 200 images where different post-processing attacks have been done to obtain a total of 10,400 images. The other 60 images have a resolution of 3000×2000 and different post-processing attacks like compression, blurring, scaling, rotation, and noise addition have been created to obtain a total of 3120 images.

3) *Coverage*: The dataset is composed of different copy-clone attacks with a total of 100 images of tampered and as well as authenticated ones. The image size is 400×486 with complex attacks like rotation, scaling, and illumination attacks.

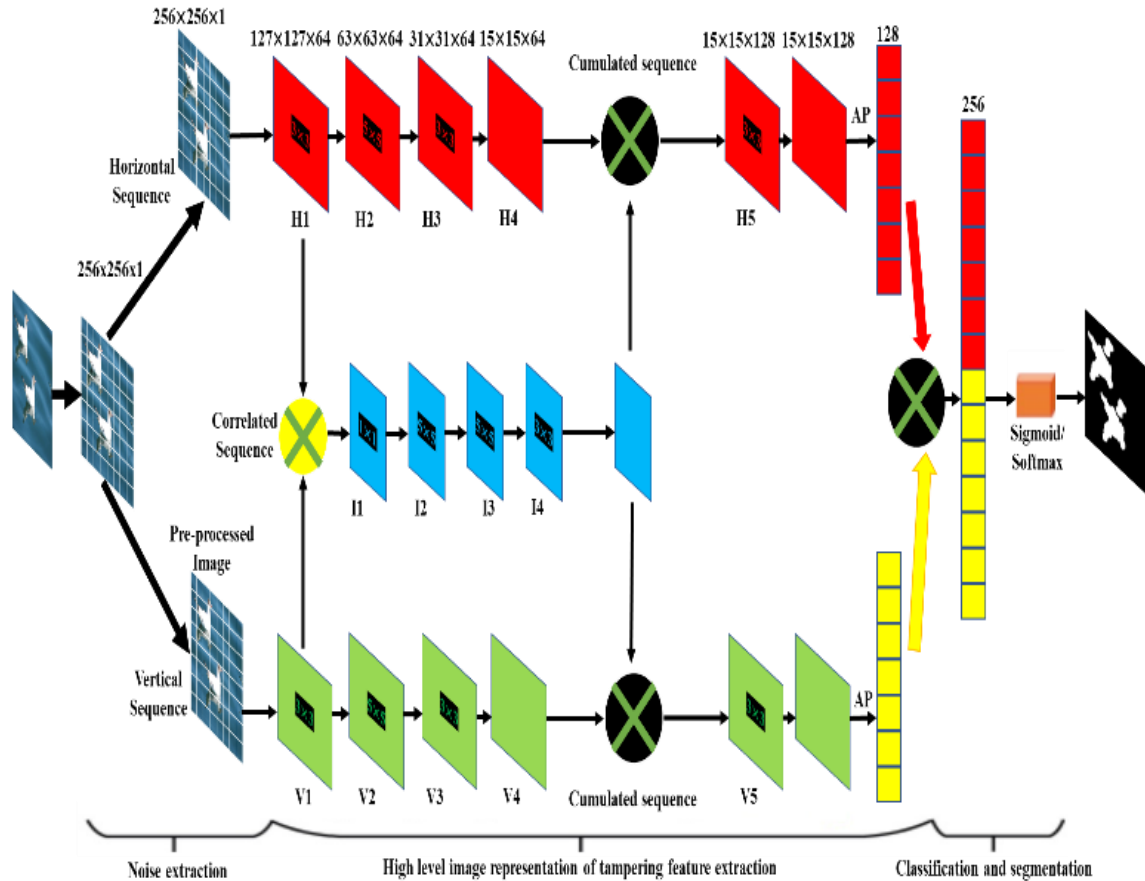


Fig. 2. Architecture of tampering detection and segmentation using improved convolution neural network.

### B. Preprocessing

The work has used a total of three datasets; in this first, the image is resized to  $512 \times 512$  into the non-overlapping region of 64 (i.e.,  $8 \times 8$ ) similar to the work presented in [17]; thus, induces certain artifacts. In [17] used space-filling curve for extracting correlation among both horizontal and vertical streams; the model achieved good, tampered region detection accuracies; however, with poor segmentation accuracies; especially under small-smooth hybrid attacks. In addressing the segmentation problem, this paper introduces an improved CNN model that with an additional layer to obtain a good correlated features-map for achieving improved tampered region detection and segmentation outcome.

## IV. PROPOSED METHODOLOGY

The methodology to localize and segment tamper regions considering hybrid attacks using TDS-ICNN is presented in this section. The feature extraction process using TDS-ICNN. Lastly, extracted features are highly correlated and training is done to create a good descriptor in classifying whether a given image is authenticated or tampered with.

### A. System Model and Architecture

The preprocessed image are passed into proposed improved CNN model for extraction of resampling features and identify the tampered region and segment it. In this work

50% of images from three different dataset is taken as input during training process of ICNN and tampering detection and segmentation model is constructed. The architecture of TDS-ICNN is given in Fig. 2. The working of tampering detection using ICNN architecture is given in Algorithm 1.

---

### Algorithm 1: the ICNN-based tampering detection and segmentation.

---

**Step 1. Start**

**Step 2. Load the images.**

**Step 3. Preprocess image into  $512 \times 512$  into the non-overlapping region of 64 (i.e.,  $8 \times 8$ )**

**Step 4. Pass the image into a three-layer ICNN.**

**Step 5. The first layer extracts the multi-dimensional RSF with the presence of noise. The RSF is captured by considering the difference between adjacent pixels across vertically and horizontally directions.**

**Step 6. The middle layer extracts the high-level feature across vertical and horizontal directions. The features that are correlated across both horizontal and vertical directions are aggregated.**

**Step 7. The, using last layer i.e., SoftMax and sigmoid function takes aggregated features as input for learning diverse features and optimizing binary tampering detection problems in multimedia forensics, respectively.**

**Step 8. Store the result and segmentation outcome.**

**Step 9. Stop**

---

A detailed explanation of the different layers is given below.

### B. Extraction of Noisy Features

In multimedia forensic extraction of resampling features is difficult as it is dependent on the information presented in a respective image. Nonetheless, some existing methodologies showed RSF extraction is not dependent on an image by extracting RSF through spatial domain using redundant feature properties. In this work, the noise is modeled by interpolating the current pixel with neighboring pixels and the difference in estimates is computed considering the image size of  $256 \times 256$ . In extracting the initial resampling feature with minimal training overhead two high-pass filters are used CNN kernel namely horizontal  $3 \times 1$  and vertical  $1 \times 3$  filters. Then the image is convoluted with padding and stride set to 1 using these filters, after that the difference (i.e., correlation) between neighboring pixels in vertical and as well horizontal direction are extracted to obtain a residual map of  $256 \times 256 \times 1$ .

### C. High-level Feature Extraction

This layer takes input from the previous layer for extraction of high-level features. The standard tampered region detection and segmentation model extracts features and correlates through each direction individually; as a result, exhibits very poor performance. However, in this paper, the RSF features are extracted and weighted in both directions individually, where it is composed of five similar groups. The group encompasses 4-layer such as convolutional layer, batch normalize layer, activation layer, and pooling layer. The fifth group has correlated features collected from the middle layer of TDS-ICNN. Finally, the features from different layers are aggregated to obtain the final RSF feature to perform tapering detection classification.

The middle layer in TDS-ICNN fuses the correlated features from both directions. The middle layer is composed of 4 groups such as convolutional layer, batch normalize layer, activation layer, and pooling layer. The feature extracted from group 1 from horizontal and vertical streams is fused considering  $1 \times 1$  convolutional kernel with stride set to 1. The remaining three groups are utilized for the extraction of high-level RSF illustrations of aggregated tampering information. Finally, by interpolating in both directions backward the feature map is established.

### D. Classification

The ICNN introduces fully-connected CNN employing SoftMax/Sigmoid operation. The model takes input features from the middle layer and performs classification based on probability estimates that belong to the tampered or non-tampered group using the following equation.

$$P(z = 1|y) = \frac{1}{1+f^{-a}} \quad (1)$$

$$P(z = k|y) = \frac{f^{a_k}}{\sum_{l=0}^L f^{a_l}} \quad (2)$$

where Eq. (1) defines the sigmoid operation of a fully connected layer for performing classification of establishing whether an image is tampered with or not as output. The parameter  $P(z = 1|y)$  defines the probability of whether  $y$  is

classified into the respective group. Eq. (2) is used for detecting multiple tampered regions using SoftMax operation, where  $a_k$  is the fully connected layer output of the  $k^{th}$  neuron. The parameter  $P(z = k|y)$  defines the probability of whether  $y$  belongs to the  $k^{th}$  group.

### E. Convolution Layer

The feature extraction done using the convolutional layer is as follows

$$G_k^{(o)} = \sum_{l=0}^L G_l^{(o-1)} * \alpha_{lk}^{(o)} + c_k^{(o)} \quad (3)$$

where  $G_k^{(o)}$  defines the  $k^{th}$  feature-map established inside the  $o^{th}$  layer,  $G_l^{(o-1)}$  represents the  $j^{th}$  feature-map established inside  $(the\ o - 1)^{th}$  layer,  $\alpha_{lk}^{(o)}$  defines the  $l$  channel of  $k^{th}$  convolutional kernel inside the  $o^{th}$  layer, and  $c_k^{(o)}$  represents  $k^{th}$  bias parameter of  $o^{th}$  layers, and  $*$  represent two-dimension convolution operation. The convolutional layer is set to 3 filters with sizes of  $(1 * 1, 3 * 3, \text{ and } 5 * 5)$  and a stride of 1.

### F. Batch Normalization

The feature map extracted in the previous layer is normalized according to feature variance according to its distribution in the middle layer. The batch normalizer operates between activation and convolutional layers. The average between total information inside the batch is described as follows

$$\beta = \frac{1}{n} \sum_{j=0}^n y_j \quad (4)$$

where  $\beta$  defines the average,  $n$  defines the overall size of the feature used, and  $y_j$  represents the  $j^{th}$  information used. In a similar, manner the difference between the total features inside the batch is estimated as follows

$$\gamma^2 = \frac{1}{n} \sum_{j=0}^n (y_j - \beta)^2 \quad (5)$$

where  $\gamma^2$  defines the difference. In this work, normalization is done on each feature to obtain new feature sets  $\hat{y}_j$  with average initialized to 0 and difference initialized to 1 and is obtained using the following equation

$$\hat{y}_j = \frac{y_j - \beta}{\sqrt{\gamma^2 + \delta}} \quad (6)$$

where  $\delta$  defines a trivial floating-point parameter higher than 0 that is used for avoiding dividing by zero error. The final batch-normalized feature is expressed as follows

$$z_j = \varphi \hat{y}_j + \omega \quad (7)$$

where,  $\varphi$  and  $\omega$  are the CNN extracted features, and  $z_j$  defines batch normalization  $j^{th}$  output. In this work to obtain better features an activation function is used that is non-linear. The adoption of such a layer will not cause significant changes due to smaller fluctuations in prediction error.

### G. Activation

In this work, the TDS is represented in the form of different spaces for achieving better-tampered region detection in multimedia forensics. The work uses TanH as an activation



function instead of ReLu and Sigmoid because it works well for features with higher differences.

#### H. Pooling Layer

The element size is reduced by down-sampling the feature maps and establishing the hierarchical structure by observing continuous features' convolutional filter. The max pooling kernel size is set to 3×3 and stride of 2 and is applied to all pooling layers except the 5<sup>th</sup> layer of both streams for providing maximum parameter in each input feature-maps by capturing patterns on neighboring pixels. The average pooling is used in the last pooling layer of both streams for down-sampling the feature maps to 1 to minimize the model parameter of fully connected CNN. The adoption of such a mechanism significantly aided in achieving improved tampered region identification and segmentation using the proposed methodology.

### V. EXPERIMENTAL STUDY

In this section experiment is done to validate the performance of TDS-ICNN over existing tampering detection methodologies like copy-move forgery detection using binary descriptor feature (CMFD-BDF) [22], BusterNet [23], fast and efficient CMFD (FE-CMFD) [24], AR-Net [26], and STRDNet [27].

#### A. Setup and Metrics

The TDS-ICNN model is modeled utilizing Python, C++, and Matlab libraries. The Intel I-7 processor with 16 GB RAM running with Windows 10 platform is used for conducting the experiments. Performance is evaluated using MICC-600, Coverage, and CoMoFoD dataset. The MICC-F600 dataset undergoes scaling and rotation post-processing tamper attacks. The CoMoFoD dataset undergoes compression, scaling, and rotation post-processing tamper attacks. The coverage dataset undergoes compression, scaling, and post-processing tamper attacks. The ROC metrics used are recall/ true positive rate (TPR), false positive rate (FPR), and F1-score for validating different tamper identification models.

$$\text{False positive rate (FPR)} = \frac{FP}{(FP + TN)} \quad (8)$$

$$\text{True positive rate (TPR)} = \frac{TP}{(TP + FN)} \quad (9)$$

$$\text{F1 - Score} = \frac{2TP}{((2TP + FN + FP))} \quad (10)$$

$$\text{Accuracy} = \frac{((TN + TP))}{((TP + FP + TN + FN))} \quad (11)$$

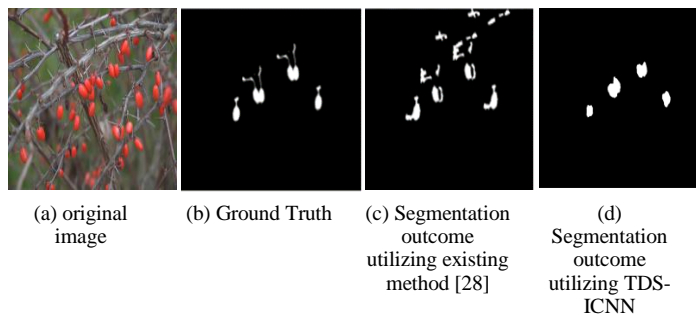


Fig. 3. Segmentation outcome of different methodologies.

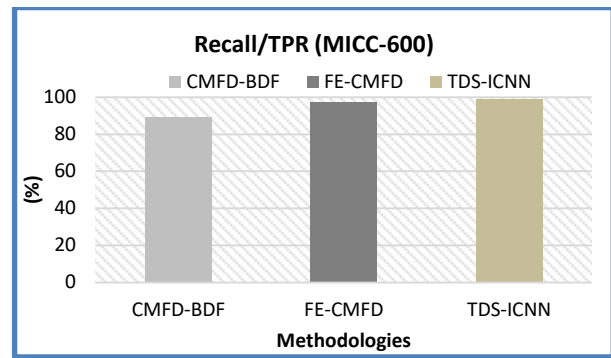


Fig. 4. Recall performance for MICC-600 dataset.

#### B. MICC Dataset

The experiment is conducted using the MICC-F600 dataset. The tampering segmentation result utilizing TDS-ICNN and other recent tamper identification models is graphically represented in Fig. 3. From Fig. 3 it is seen that TDS-ICNN provides improved tampering region segmentation outcomes when compared with existing models. Fig. 4 shows recall performance achieved utilizing TDS-ICNN and other existing tampering detection methodologies. Fig. 5 shows false positive rate performance achieved utilizing TDS-ICNN and other existing tampering detection methodologies. Fig. 6 shows the F1-score at image level performance achieved utilizing TDS-ICNN and other existing tampering detection methodologies. Fig. 7 shows that the F1-score at pixel-level performance was achieved utilizing TDS-ICNN and other existing tampering detection methodologies. The outcome obtained from Table I shows that TDS-ICNN improves detection accuracy and reduces false positives; thus, it can be adapted to provide a reliable tamper identification model.

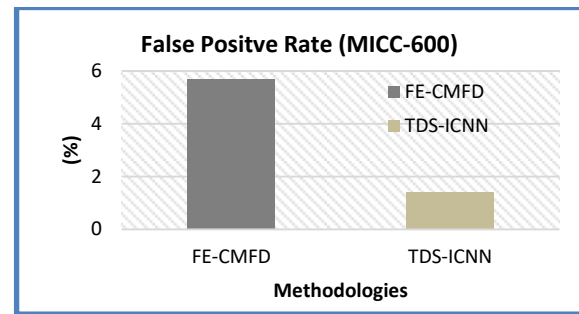


Fig. 5. False positive rate for MICC-600 dataset.

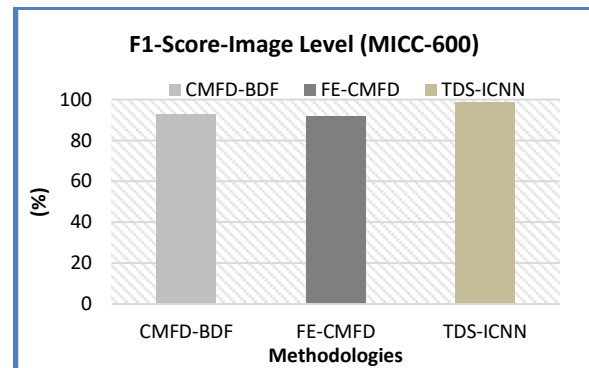


Fig. 6. F1-Score at image level performance for MICC-600 dataset.

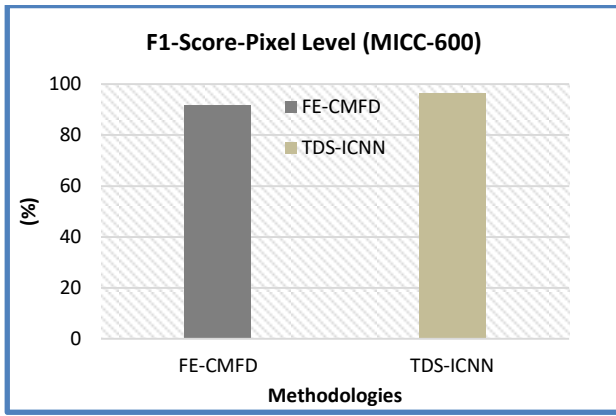


Fig. 7. F1-Score at pixel level performance for MICC-600 dataset.

TABLE I. COMPARATIVE STUDY FOR MICC DATASET

Methodology used	Performance metrics			
	TPR	FPR	F1-Score image	F1-Score pixel
CMFD-BDF [22]	89.14		92.6	
FE-CMFD [24]		5.68	91.5	91.8
TDS-ICNN [Proposed]	99.1	1.4	98.6	96.5

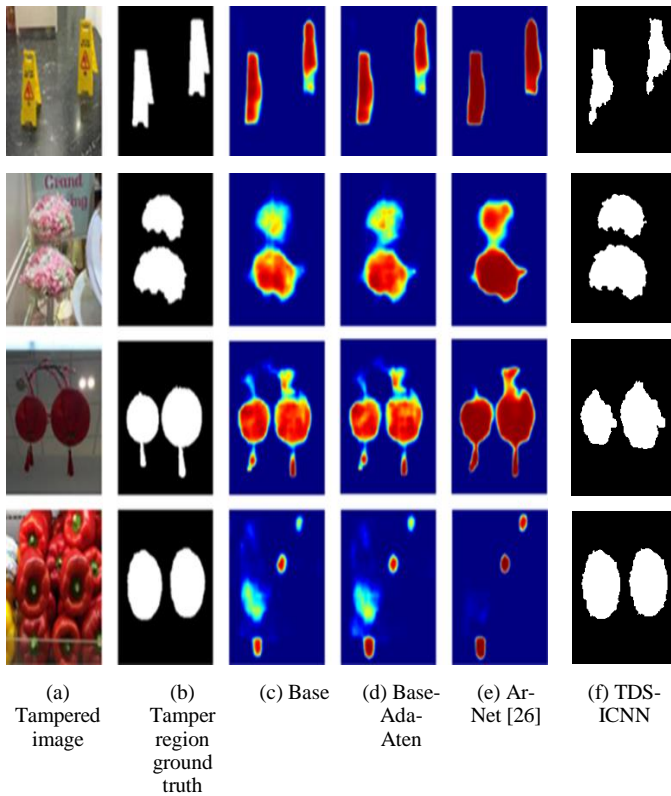


Fig. 8. Tampering region segmentation outcome using coverage dataset of proposed tampering and existing AR-Net tampering detection method.

### C. Coverage Dataset

Here experiment is carried out using a coverage dataset. In the dataset is very difficult to classify which is authenticated and which is tampered one. The tampering segmentation

results utilizing TDS-ICNN and other recent tamper identification models are graphically represented in Fig. 8 and Fig. 9. The result proves improved tamper region segmentation outcomes utilizing TDS-ICNN concerning recent tamper identification models. Fig. 10 shows the accuracy of performance achieved utilizing TDS-ICNN and other existing tampering detection methodologies. Fig. 11 shows the F1-score utilizing TDS-ICNN and other recent tamper identification methodologies. The outcome obtained from Table II shows that TDS-ICNN improves detection accuracy and reduce false positive and hence, it can be adopted to provide a reliable tamper identification model.

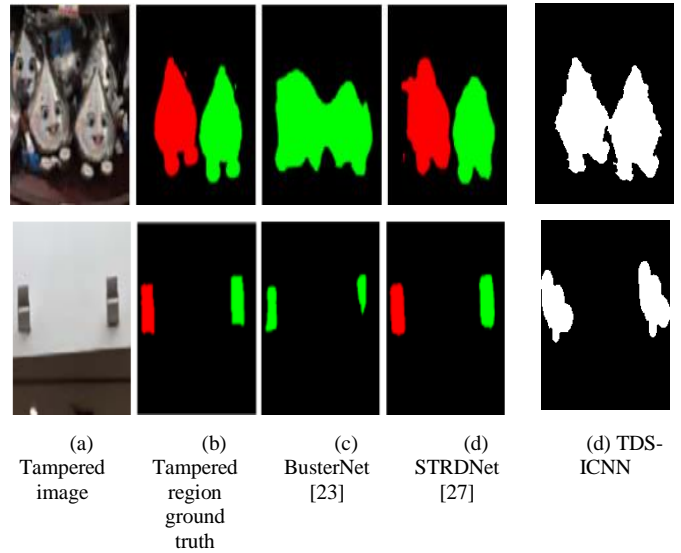


Fig. 9. Tampering region segmentation outcome using Coverage dataset of proposed tampering and existing STRDNet tampering detection method.

TABLE II. COMPARATIVE STUDY FOR COVERAGE DATASET

Methodology used	Performance metrics	
	Accuracy	F1-Score
Base [26]	0.8581	
Base-Ada-Atten [26]	0.8542	
AR-Net [26]	0.8488	
BusterNet [27]		0.464
STRDNet [27]		0.677
TDS-ICNN [Proposed]	0.8563	0.7456

### D. CoMoFoD Dataset

The CoMoFoD dataset is utilized for studying the performance of TDS-ICNN with other recent tamper identification models. The dataset has diverse post-processing attacks being accrued out; thus, making it extremely challenging to detect tamper regions and localize them. The tampering segmentation result utilizing TDS-ICNN and other recent tamper identification models is graphically represented in Fig. 12. From Fig. 12 it can be stated that TDS-ICNN improves tamper region segmentation outcomes when compared with existing models. Fig. 13 shows recall performance achieved utilizing TDS-ICNN and other existing

tampering detection methodologies. Fig. 14 shows the precision performance achieved utilizing TDS-ICNN and other existing tampering detection methodologies. Fig. 15 shows the F1-score result utilizing TDS-ICNN and other recent tamper identification methodologies. The outcome obtained in Table III shows that TDS-ICNN improves detection accuracy and reduces false positives; thus, can be adapted to provide a reliable tamper identification model.

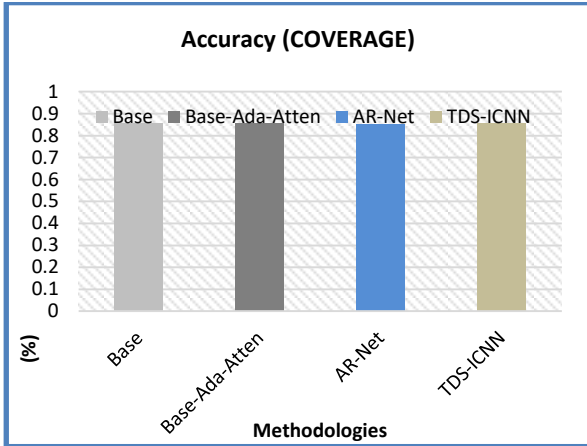


Fig. 10. Accuracy performance for coverage dataset.

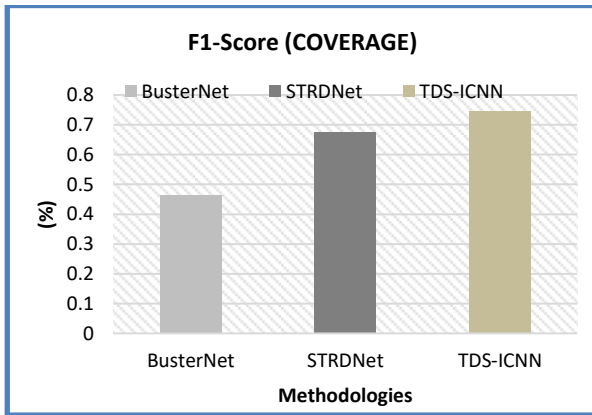


Fig. 11. F1-Score performance for coverage dataset.

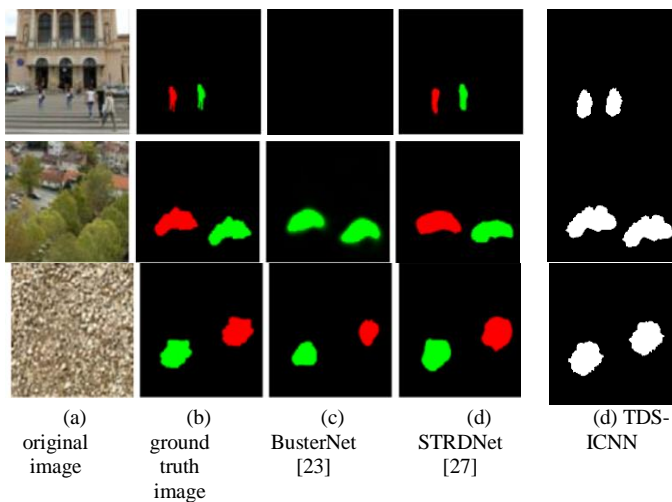


Fig. 12. Tampering region segmentation outcome using CoMoFoD dataset.

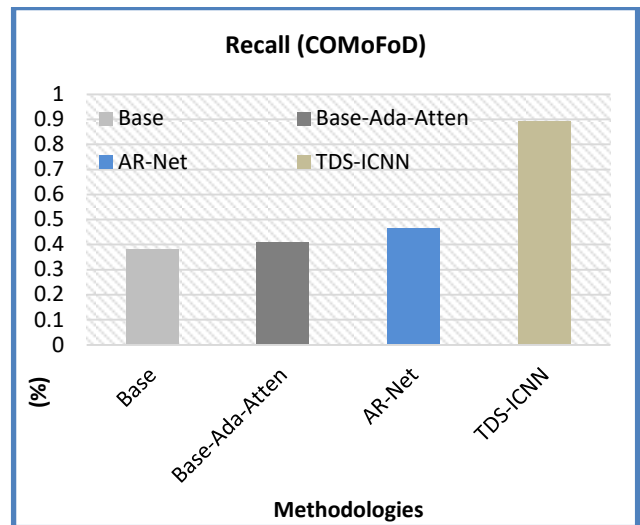


Fig. 13. Recall performance for the CoMoFoD dataset.

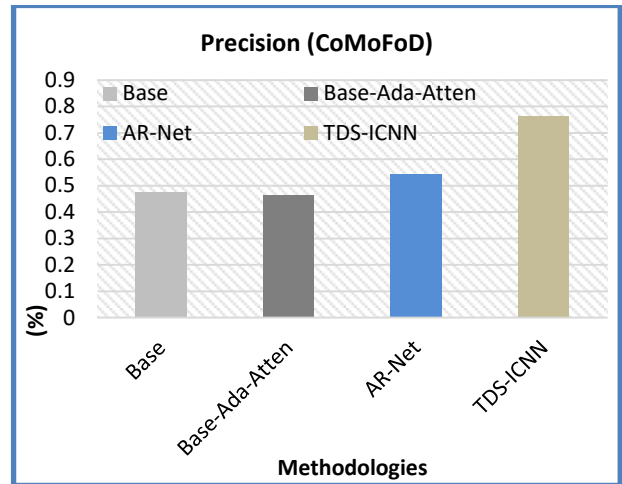


Fig. 14. Precision performance for CoMoFoD dataset.

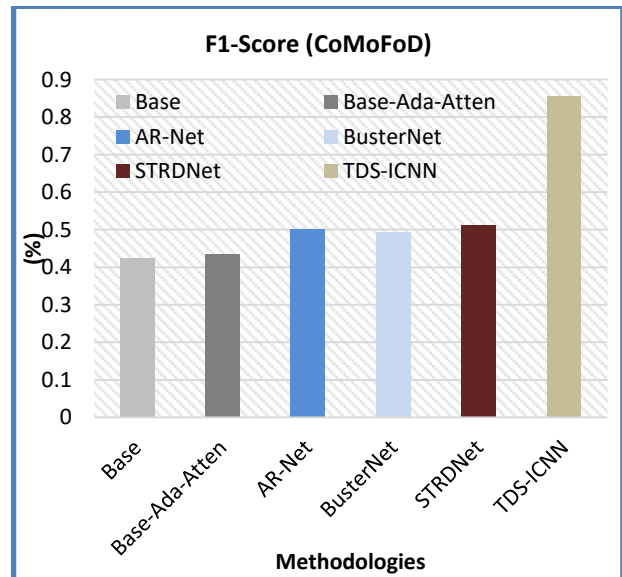


Fig. 15. F1-Score performance for CoMoFoD dataset.

TABLE III. COMPARATIVE STUDY FOR CoMoFoD DATASET

Methodology used	Performance metrics		
	Recall	Precision	F1-Score
Base [26]	0.3811	0.4768	0.4236
Base-Ada-Atten [26]	0.4075	0.4661	0.4349
AR-Net [26]	0.4655	0.5421	0.5009
BusterNet [27]	x	x	0.493
STRDNet [27]	x	x	0.511
TDS-ICNN [Proposed]	0.89	0.7654	0.856

## VI. CONCLUSION

The research work has presented a technique namely TDS-ICNN to identify whether an image is authenticated or tampered with. The preprocessing technique and feature extraction technique adopted in TDS-ICNN can retain spatial features concerning different patches. Alongside this, a good correlation exists among both horizontal and vertical curves through the introduction of a correlation layer. To eliminate spatial dependencies, the features extracted are aggregated and a descriptor is constructed to perform classification. The experiment is conducted using three datasets, such as MICC-600, Coverage, and CoMoFoD. For the MICC dataset the existing methods namely CMFD-BDF attains a TPR and F1-Score of 89.14% and 92.6%, respectively; however, the proposed TDS-ICNN attains a TPR and F1-score of 99.1% and 98.6%, respectively. For the Coverage dataset the existing methods namely AR-Net attain an accuracy of 84.88% and the proposed TDS-ICNN attains an accuracy of 85.63%, respectively. Similarly, the STRDNet attains an F1-score of 67.7%, and the proposed TDS-ICNN attains an F1-Score of 74.56%. For the CoMoFoD dataset the existing methods namely AR-Net attains a recall, precision, and F1-Score of 46.55%, 54.21%, and 50.09%, respectively; however, the proposed TDS-ICNN attains a recall, precision, and F1-Score of 89.0%, 76.54%, and 85.6%, respectively. The result attained shows that superior performance is achieved using TDS-ICNN in comparison with other standard tamper detection methods. A good ROC performance such as TPR, FPR, F1-Score, and accuracy in comparison with other existing tamper detection methodologies is achieved. The significant result provides a satisfactory benchmark for using it for real-time tampering image circulation in social media platforms and WhatsApp messenger; thereby can prevent misleading information circulation.

Future work would be focused on studying the model performance on other standard datasets like CASIA, and DO. The work would further investigate how the proposed model can be used to detect tampering in video. Further, would focus on developing an ensemble learning model to improve tampering detection accuracy with fewer false positives.

## REFERENCES

[1] Dadkhah, S., Mazzola, G., Uliyan, D., Sadeghi, S., Jalab, H.A.: State of the art in passive digital image forgery detection: copy-move image forgery. *Pattern Anal. Appl.* 21, 291–306, 2017.

[2] H. Li, W. Luo, X. Qiu and J. Huang, "Image Forgery Localization via Integrating Tampering Possibility Maps," in *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 5, pp. 1240-1252, 2017. DOI: 10.1109/TIFS.2015.2423261.

[3] J. H. Bappy, A. K. Roy-Chowdhury, J. Bunk, L. Nataraj, and B. Manjunath. Exploiting spatial structure for localizing manipulated image regions. In *ICCV*, 2017.

[4] J. Bunk, J. H. Bappy, T. M. Mohammed, L. Nataraj, A. Flenner, B. Manjunath, S. Chandrasekaran, A. K. Roy-Chowdhury, and L. Peterson. Detection and localization of image forgeries using resampling features and Deep Learning. In *Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017 IEEE Conference on, pages 1881–1889, 2017.

[5] W. Wang, J. Dong, and T. Tan. Exploring DCT coefficient quantization effects for local tampering detection. *IEEE Transactions on Information Forensics and Security*, 9(10):1653–1666, 2014. DOI: 10.1109/TIFS.2014.2345479.

[6] Yang, Hong-Ying & Qi, Shu-Ren & Niu, Ying & Niu, Pan-Pan & Wang, xiang yang. (2019). Copy-move forgery detection based on adaptive keypoints extraction and matching. *Multimedia Tools and Applications*. 78. 10.1007/s11042-019-08169-w.

[7] J. H. Bappy, C. Simons, L. Nataraj, B. S. Manjunath and A. K. Roy-Chowdhury, "Hybrid LSTM and Encoder–Decoder Architecture for Detection of Image Forgeries," in *IEEE Transactions on Image Processing*, vol. 28, no. 7, pp. 3286-3300, July 2019, doi: 10.1109/TIP.2019.2895466.

[8] Ali, S.S.; Ganapathi, I.I.; Vu, N.-S.; Ali, S.D.; Saxena, N.; Werghi, N. Image Forgery Detection Using Deep Learning by Recompressing Images. *Electronics* 2022, 11, 403. <https://doi.org/10.3390/electronics11030403>.

[9] Qazi, E.U.H.; Zia, T.; Almorjan, A. Deep Learning-Based Digital Image Forgery Detection System. *Appl. Sci.* 2022, 12, 2851. <https://doi.org/10.3390/app12062851>.

[10] Shivanandappa, Manjunath & Patil, Malini. Extraction of image resampling using correlation aware convolution neural networks for image tampering detection. *International Journal of Electrical and Computer Engineering*. 12. 3033. 2022, <https://doi.org/10.11591/ijece.v12i3.pp3033-3043>.

[11] H. Wu, J. Zhou, J. Tian, J. Liu and Y. Qiao, "Robust Image Forgery Detection Against Transmission Over Online Social Networks," in *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 443-456, 2022, doi: 10.1109/TIFS.2022.3144878.

[12] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

[13] S. Manjunatha. and M. M. Patil, "Deep learning-based Technique for Image Tamper Detection," 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), 2021, pp. 1278-1285, doi: 10.1109/ICICV50876.2021.9388471.

[14] Wang, Chengyou & Zhang, Zhi & Zhou, Xiao. (2018). An image copy-move forgery detection scheme based on A-KAZE and SURF features. *Symmetry*. 10. 706. 10.3390/sym10120706.

[15] Liang Y, Fang Y, Luo S and Chen B. Image Resampling Detection Based on Convolutional Neural Network. 2019 15th International Conference on Computational Intelligence and Security (CIS), Macao, China, 2019. pp. 257-261. DOI: 10.1109/CIS.2019.00061

[16] Shivanandappa, Manjunath & Patil, Malini. Efficient resampling features and convolution neural network model for image forgery detection. *Indonesian Journal of Electrical Engineering and Computer Science*. 25. 183, 2022. <https://doi.org/10.11591/ijeecs.v25.i1.pp183-190>.

[17] Shivanandappa, Manjunath & Patil, Malini. Tampering Detection using Resampling Features and Convolution Neural Networks. *Turkish Journal of Computer and Mathematics Education; Trabzon Vol. 12, Iss. 11*, pp. 2791-2800, 2021.

[18] Zhou, G., Tian, X. & Zhou, A. Image copy-move forgery passive detection based on improved PCNN and self-selected sub-images. *Front.*

- Comput. Sci. 16, 164705 (2022). <https://doi.org/10.1007/s11704-021-0450-5>.
- [19] Flenner, Arjuna & Peterson, Lawrence & Bunk, Jason & Mohammed, Tajuddin Manhar & Nataraj, Lakshmanan & Manjunath, B. Resampling Forgery Detection Using Deep Learning and A-Contrario Analysis. *Electronic Imaging*. 2018. 10.2352/ISSN.2470-1173.2018.07.MWSF-212, 2018.
- [20] Qazi, Tanzeela & Ali, Mushtaq & Hayat, Khizar & Baptiste, Magnier. (2022). Seamless Copy–Move Replication in Digital Images. *Journal of Imaging*. 8. 69. 10.3390/jimaging8030069.
- [21] Huang, H., Ciou, A. Copy-move forgery detection for image forensics using the superpixel segmentation and the Helmert transformation. *J Image Video Proc.* 2019, 68 (2019). <https://doi.org/10.1186/s13640-019-0469-9>, 2019.
- [22] Raju, P.M., Nair, M.S.: Copy-move forgery detection using binary discriminant features. *J. King Saud Univ. - Comput. Inf. Sci.* 2018.
- [23] Yue Wu, Wael Abd-Almageed, Prem Natarajan. BusterNet: Detecting Copy-Move Image Forgery with Source/Target Localization. *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 168-184.
- [24] Y. Li and J. Zhou, "Fast and Effective Image Copy-Move Forgery Detection via Hierarchical Feature Point Matching," in *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 5, pp. 1307-1322, May 2019, doi: 10.1109/TIFS.2018.2876837.
- [25] H. Chen, X. Yang and Y. Lyu, "Copy-Move Forgery Detection Based on Keypoint Clustering and Similar Neighborhood Search Algorithm," in *IEEE Access*, vol. 8, pp. 36863-36875, 2020, doi: 10.1109/ACCESS.2020.2974804.
- [26] Y. Zhu, C. Chen, G. Yan, Y. Guo and Y. Dong, "AR-Net: Adaptive Attention and Residual Refinement Network for Copy-Move Forgery Detection," in *IEEE Transactions on Industrial Informatics*, vol. 16, no. 10, pp. 6714-6723, Oct. 2020, doi: 10.1109/TII.2020.2982705.
- [27] B. Chen, W. Tan, G. Coatrieux, Y. Zheng and Y. Q. Shi, "A serial image copy-move forgery localization scheme with source/target distinguishment," in *IEEE Transactions on Multimedia*, 2020. doi: 10.1109/TMM.2020.3026868.
- [28] J. Li, X. Li, B. Yang, and X. Sun. Segmentation-based image copy-move forgery detection scheme. *IEEE Transactions on Information Forensics and Security*, 10(3):507–518, 2015.
- [29] Li Q, Wang C, Zhou X, Qin Z. Image copy-move forgery detection and localization based on super-BPD segmentation and DCNN. *Sci Rep.* 2022 Sep 2;12(1):14987. doi: 10.1038/s41598-022-19325-y.
- [30] Fu, G.; Zhang, Y.; Wang, Y. Image Copy-Move Forgery Detection Based on Fused Features and Density Clustering. *Appl. Sci.* 2023, 13, 7528. <https://doi.org/10.3390/app13137528>.
- [31] Manaf Mohammed Ali Alhaidery, Amir Hossein Taherinia, Haider Ismael Shahadi, A robust detection and localization technique for copy-move forgery in digital images, *Journal of King Saud University - Computer and Information Sciences*, Volume 35, Issue 1, 2023, Pages 449-461, <https://doi.org/10.1016/j.jksuci.2022.12.014>.

# PRESSNet: Assessment of Building Damage Caused by the Earthquake

Dewa Ayu Defina Audrey Nathania<sup>1</sup>, Alexander Agung Santoso Gunawan<sup>2</sup>, Edy Irwansyah<sup>3</sup>  
Computer Science Department, Master of Computer Science, Bina Nusantara University, Jakarta, Indonesia<sup>1</sup>  
Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia<sup>2,3</sup>

**Abstract**—Loss of life and property often occur due to natural disasters and other significant occurrences like earthquakes, which make manual damage assessment a time-consuming and inefficient process. In an attempt to address this challenge, researchers have been investigating the field of automated damage assessment in Remote Sensing. With time, this area of research has transformed from conventional machine learning techniques to more sophisticated deep learning techniques. The study puts forward the PRESSNet model as a solution for assessing building damage. The effectiveness of the proposed PRESSNet model is compared to that of a baseline model, PSPNet, and ResNet 50, across different types of damage. This study contributes by introducing the spatial attention module to the baseline model. The xBD Dataset was used both before and after the Palu earthquake disaster. The results show that PRESSNet performs similarly or slightly better than the baseline model in all damage categories. This illustrates the impressive ability of the proposed PRESSNet architecture to accurately detect and classify building damage. This research sheds light on the development of effective models for assessing disaster damage and lays the foundation for future progress in this crucial area.

**Keywords**—Remote sensing; deep learning; PSPNet; ResNet; spatial attention

## I. INTRODUCTION

In an era characterized by an increasing frequency of natural disasters, including earthquakes, floods, and hurricanes, which yield dire consequences, the importance of proficient crisis management becomes paramount. These catastrophic events result in not only the unfortunate loss of human lives but also substantial property damage [1]. Access to crucial information, both before and after a catastrophic event, proves to be of utmost importance in enhancing disaster response strategies and mitigating the impact on human lives and physical infrastructure [2]. In addition to enhancing the capacity for early detection and warning before the crisis begins, it is crucial to gather information about a disaster as soon as it occurs [3].

Assessing structural damage to buildings stands as a critical issue in the field of disaster response, as it has been identified as a prominent factor contributing to the loss of life during natural calamities [4]. The precise assessment of such harm is crucial to enhance emergency response efforts and ultimately preserve a greater number of lives. Recent advancements in remote sensing technology and the deployment of satellite constellations have greatly enhanced our ability to obtain high-resolution satellite (HRS) data [5]. Integrating this wealth of data with machine learning (ML) and deep learning (DL)

methodologies offers a promising approach for evaluating structural damage in the aftermath of a calamitous event [5].

Assessing structural damage to buildings stands as a critical issue in the field of disaster response, as it has been identified as a prominent factor contributing to the loss of life during natural calamities [4]. The precise assessment of such harm is crucial to enhance emergency response efforts and ultimately preserve a greater number of lives. Recent advancements in remote sensing technology and the deployment of satellite constellations have greatly enhanced our ability to obtain high-resolution satellite (HRS) data. Integrating this wealth of data with machine learning (ML) and deep learning (DL) methodologies offers a promising approach for evaluating structural damage in the aftermath of a calamitous event [5].

However, recent evaluations of the PSPNet model, which employs ResNet 50 as its underlying framework, have identified shortcomings in effectively gathering and leveraging contextual information at various scales, particularly in the intricate realm of building degradation [8]. To address this challenge, the present study introduces the PRESSNet framework, which utilizes spatial attention mechanisms to enhance the identification and prioritization of critical regions inside a building that exhibit signs of structural deterioration. PRESSNet's performance in accurately evaluating the extent of building damage surpasses that of the PSPNet + ResNet-50 model, showcasing its capacity to augment disaster response efforts.

The primary objective of this study is to underscore the importance of spatial attention mechanisms within deep learning models for the purpose of disaster response. PRESSNet's contribution to the advancement of computer vision research in disaster management lies in its emphasis on the effective identification and classification of different levels of damage. This study highlights the significance of attention mechanisms in improving the performance of convolutional neural networks, thereby contributing to the development of more efficient disaster response systems.

Furthermore, this study introduces the PRESSNet model as a solution for assessing building damage. It compares the effectiveness of PRESSNet to that of a baseline model, PSPNet, and ResNet 50, across different types of damage. The xBD Dataset is utilized both before and after the Palu earthquake disaster. The results demonstrate that PRESSNet performs similarly or slightly better than the baseline model in all damage categories. The baseline model exhibits strong performance with a macro-average F1 score of 89.5%, while

PRESSNet slightly outperforms it, achieving a macro-average F1 score of 90%. This illustrates the impressive ability of the proposed PRESSNet architecture to accurately detect and classify building damage. This research sheds light on the development of effective models for assessing disaster damage and lays the foundation for future progress in this crucial area.

## II. RELATED STUDIES

### A. Building Damage Assessment in High Resolution Satellite Imagery using Deep Learning

The assessment of building damage using deep learning models, particularly convolutional neural networks (CNNs), has been a prominent subject of investigation within the domain of remote sensing utilizing satellite and aerial photography. The authors, Garg et al. [8], introduced a deep convolutional neural network (CNN) that utilizes transfer learning to do building damage assessment using satellite photos. The model underwent training by utilizing a pre-existing VGG16 network, which was subsequently fine-tuned using a dataset specifically curated for earthquake-affected buildings. The results indicate that the suggested model had superior performance compared to standard machine learning techniques, achieving an accuracy of 87% on the test dataset.

The authors, Hu et al. [9], introduced an innovative methodology that uses deep learning techniques to evaluate structural harm in buildings through the analysis of aerial photography. The employed methodology involved the utilization of a Siamese network architecture to conduct a comparative analysis of aerial photos captured before and after a catastrophic event. The objective was to accurately detect and pinpoint areas where structural damage to buildings had occurred. The model achieved an overall accuracy of 92.65 percent on the test dataset.

Using satellite data, Shah et al. [10] created a deep convolutional neural network (CNN) method that makes use of transfer learning to assess the degree of building damage. The model under consideration utilized a pre-existing ResNet-50 network [11], which was subsequently refined through the process of fine-tuning. A dataset of buildings damaged by Hurricane Harvey served as the basis for this refinement. Waseem et al. [12] developed a system based on deep learning methods to assess the degree of building damage using satellite data. The proposed model integrates features derived from a pre-trained ResNet 50 network with manually engineered characteristics, including texture and color features.

### B. PSPNet

The assessment of building damage holds significant importance in the aftermath of disasters, as it serves to guide relief operations, optimize resource allocation, and ascertain the overall consequences of both natural and human-induced calamities. The assessment of building damage has been significantly improved with the application of convolutional neural networks (CNNs), thanks to recent breakthroughs in computer vision and deep learning. The Pyramid Scene Parsing Network (PSPNet) is a convolutional neural network (CNN) architecture that has gained significant recognition due to its capacity to effectively capture contextual information at several scales.

The Pyramid Scene Parsing Network (PSPNet) is a convolutional neural network that uses a pyramid pooling module to capture contextual information at many scales. Additionally, it incorporates a spatial attention module to enhance important features and suppress non-informative ones [13]. The inclusion of the pyramid pooling module in PSPNet facilitates the efficient extraction of features at different scales. This capability is crucial in accurately identifying zones of building damage that exhibit diverse sizes. The architecture of PSPNet incorporates a pyramid pooling module that effectively combines contextual information from multiple scales. The module presented in this study is designed to gather contextual information at many scales. This enables the model to effectively capture both global and local contextual information that is relevant for building damage assessment. The inclusion of the pyramid pooling module significantly contributes to the model's ability to effectively differentiate between regions that are damaged and those that are unaffected.

Dilated convolutions, which are also called atrous convolutions, are used in the PSPNet architecture to increase the receptive field while keeping the spatial resolution the same. Dilated convolutions enable the neural network to record a wider contextual range by introducing gaps inside the convolutional kernels, thereby maintaining the integrity of spatial details. The PSPNet employs a fusion technique to effectively integrate contextual information by merging elements at many scales. The final segmentation map is generated by combining and refining the multi-scale feature maps from the pyramid pooling module using convolutional layers.

The PSPNet model comprises three primary components, each serving distinct roles. By using ResNet 50 as the underlying framework, it is possible to create a feature map from an input image. The pyramid parsing module utilizes the feature map of the initial module, ResNet 50, to extract representations of four separate sub-representations. These sub-representations are obtained using convolution operations of sizes  $1 \times 1$ ,  $2 \times 2$ ,  $3 \times 3$ , and  $6 \times 6$ . The result of the representation that was restored in the preceding layer is increased in size and combined with all of the increased representations, together with the characteristics that were mapped in the initial module. The convolutional layer utilizes a representation of the input image produced from the second module in order to obtain the final semantic segmentation results.

The point-wise attention blinders used for the research conducted by Hamdi et al. [14] exhibit dissimilarities when compared to the blinders employed in other studies. The PSA module specifically instructs on the implementation of location- and category-sensitive masks that possess self-adaptive capabilities. The Public Service Announcement (PSA) acquires the ability to gather contextual information for each unique point in a manner that is flexible and tailored to the specific needs of the user.

The method of pooling known as shift pooling was introduced by Yuan et al. [15] and was implemented to enhance the performance of PSPNet. The repositioning of the pooling grid allows for the comprehensive acquisition of local

feature information by the pixels located at the edges and corners of the grid, resulting in enhanced segmentation outcomes.

### C. Spatial Attention Mechanism Module

The spatial attention mechanism (SAM) has been identified as a highly successful approach for evaluating building damage and other computer vision tasks. The Spatial Attention Module (SAM) is employed to consolidate comparable picture information and augment the network's capacity to depict these characteristics. The utilization of SAM allows the model to effectively allocate importance to informative traits while simultaneously suppressing non-informative ones through the selective concentration on pertinent regions. This approach has been widely employed in several deep learning models to evaluate the extent of building damage.

Chen et al. [16] introduced a novel approach for enhancing the precision and consistency of building damage assessment. Their proposed method involves utilizing a change detection feature extractor, which incorporates a pyramid spatial temporal attention module. The experiments showed that this module, called SAM, makes it possible for the network to capture similar features and highlight the unique features of damaged regions.

The domain of remote sensing through satellite and aerial imaging has witnessed significant research activity in the realm of building damage assessment. Deep learning models, particularly convolutional neural networks (CNNs), have emerged as a prominent approach in this subject. The authors, Garg et al. [8], introduced a deep convolutional neural network (CNN) that utilizes transfer learning to do building damage assessment using satellite photos. The model underwent training by utilizing a pre-existing VGG16 network and subsequently underwent refinement through the utilization of a dataset consisting of buildings impacted by earthquakes. The results indicate that the suggested model exhibited superior performance compared to standard machine learning techniques, achieving an accuracy rate of 87% on the test dataset.

Attention mechanisms are widely used in deep learning. An attention model was created by Mnih et al. [17] that picks a number of regions or sites for processing in an adaptive manner. Multiple attention masks were found by Chen et al. [18] to combine feature maps or forecasts from many branches. Pre- and Post-Disaster Imagery were recovered from a single model with the same weight in investigations conducted by Weber et al. [19], and the output features were layered (concatenated) to derive features between pre and post.

A self-attention machine translation model was developed by Vaswani et al. [20]. The correlation matrix between each spatial location in the feature map was calculated by Wang et al. [21] to identify attention masks. Zhao et al.'s [22] point-wise spatial attention network (PSANet) was suggested as a way to relax the local neighborhood constraint. A self-adaptively learnt attention mask connects each location on the feature map to every other location. Additionally, scene understanding is allowed through bidirectional information propagation. The

forecast of the current position can be aided by knowledge from previous positions, and the prediction of other positions can benefit from information from the current one.

## III. RESEARCH METHODOLOGY

### A. Dataset

The applied dataset is the 2018 Palu earthquake from the Tier 1 XBD dataset (<https://xview2.org>). Fig. 1. shows the map of Palu. No damage, minor damage, major damage, and destroyed are the four levels of damage. Pre- and post-disaster dataset images are distributed as follows: 80% to the training dataset and 20% to the test dataset.

### B. Data Preprocessing

The original dimension of the Palu dataset was 128 x 128 pixels, which has been reduced to 64 x 64. No Damage and Destroyed are the two classes or levels within this research. Pre- and post-disaster images for the Train Dataset increased from 54 to 3264. Pre and Post Images increased from 15 Images (1024x1024) to 1024 Images (256x256) in the test dataset. The training dataset is composed of 70% Training data and 30% Validation data. The training dataset is then rotated by 30 degrees to generate a new patch. The Steps per Epoch used for this research are 16000 for Training and 5000 for validation.

The training dataset was used to optimize the neural network, whereas the validation dataset was used to evaluate the network's training efficacy. Using the evaluation dataset, the effectiveness of the optimized neural network was determined. The validation and training images were read into (256, 256, 4) arrays of the form. The labels were interpreted as an array of shapes (256, 256, 1), with the values 1 or 0 indicating whether the item is damaged or not.

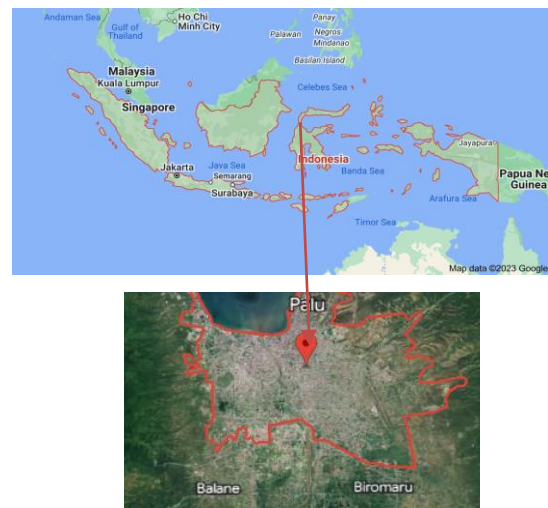


Fig. 1. Palu earthquake, 2018.

### C. Proposed Method

This study proposed the model PRESS-Net, which consists of the single segmentation model PSPNet[10] using ResNet[12] as a with different backbone using ResNet 50 and ResNet 101 with Hyperparameter Tuning and Spatial Attention



Module [23] to learn which feature and where the location of feature is important to be selected.

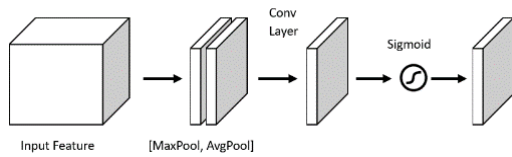


Fig. 2. Spatial attention module architecture.

The spatial attention map is generated by applying the interspatial relationship between features. When compared to channel attention, the spatial attention module architecture (see Fig. 2) emphasizes "where" as an additional information component. We first use average-pooling and max-pooling operations along the channel axis to compute spatial attention, and we then concatenate the results to provide an efficient feature descriptor. It has been proven that aggregating procedures applied along the channel axis can effectively highlight informative regions. We apply a convolution layer to the concatenated feature descriptor to generate a spatial attention map  $M_s(F) \in R^{H \times W}$  that encodes where to highlight or suppress.

The computation for spatial attention is as follows:

$$M_s(F) = \sigma(f^{7 \times 7}([AvgPool(F); MaxPool(F)])) \\ = \sigma(f^{7 \times 7}([F_{avg}^s; F_{max}^s]))$$

- Where  $\sigma$  denotes the sigmoid function and  $f^{7 \times 7}$  represents a convolution operation with the filter size of  $7 \times 7$  [23].

---

**Algorithm 1: Pseudocode for building segmentation using CNNs**

**Inputs:** xBD Dataset

**Output:** Segmented building damage

1. **Start:**
2. Load and preprocess the xBD Dataset
3. Apply noise removal to the Palu dataset
4. Initialize training loop counter  $i = 1$
5. Set maximum training loops as MAX\_LOOPS
6. Do while  $i \leq MAX\_LOOPS$ :
7. # Training Data Augmentation
8. Augment training data and masks
9. # Split dataset into training and validation sets
10. Split Augmented Data into  $X_{train}$ ,  $X_{val}$  and Augmented Masks into  $y_{train}$ ,  $y_{val}$
11. # Create the PSPNet model with ResNet-50 backbone and spatial attention
12.  $pspnet\_model = create\_pspnet(input\_shape, num\_classes)$
13. # Train the model
14. Train  $pspnet\_model$  on  $X_{train}$  and  $y_{train}$  for a specified number of epochs
15. # Evaluate the model's performance on validation set
16. Evaluate  $pspnet\_model$  on  $X_{val}$  and  $y_{val}$  to get validation loss and accuracy
17. # Print evaluation results (optional)
18. Print "Validation Loss:", validation\_loss
19. Print "Validation Accuracy:", validation\_accuracy
20. # Save the trained model (optional)
21. Save  $pspnet\_model$  to disk with a suitable filename
22. # Increment training loop counter
23. Increment  $i$

24. End do while
25. End

- The process begins by loading a dataset called xBD, containing images of building damage. These images are prepared for analysis by removing any distracting noise from them, which helps the algorithm focus on the key information. To train an accurate model, a training loop is set up. A loop counter named 'i' starts at 1, and a maximum number of loops (MAX\_LOOPS) is predetermined to manage the training process. Inside the loop, data augmentation is applied to the training images and their corresponding masks.
- This augmentation involves making small changes to the images, like flipping or rotating, to create a more varied dataset. This diversity helps the model learn better. The augmented data is divided into two parts: one for training ( $X_{train}$  and  $y_{train}$ ) and another for validation ( $X_{val}$  and  $y_{val}$ ). This separation helps evaluate how well the model performs on new, unseen data. Within the loop, a specialized model called PSPNet is constructed. It's designed to understand and segment building damage.
- This model uses a ResNet-50 backbone, which helps identify important features, and a spatial attention module to focus on critical parts of the image. The PSPNet model is trained using augmented training data. It learns from the images and their associated masks that indicate where building damage is present. This training process continues for a set number of cycles, improving the model's accuracy with each iteration. After training, the model's performance is tested on the validation dataset. This helps measure how well the model has learned to identify building damage. The results, such as validation loss and accuracy, can be printed out to assess the model's progress. If desired, the trained PSPNet model can be saved to the computer's storage. This way, the model can be reused later without needing to go through the training process again. The loop repeats as long as the loop counter 'i' is within the set maximum number of loops (MAX\_LOOPS). In each loop, the model's understanding of building damage is refined, leading to better performance.
- Once the desired number of loops is completed, the algorithm finishes its execution. This systematic approach helps create a reliable model for segmenting building damage from images, contributing to more accurate and efficient analyses.
- PRESSNet is a deep learning model that incorporates the PSPNet model, ResNet 50 architecture, and a spatial attention mechanism to improve image segmentation performance. The evaluation metrics for this research will be the macro-average F1 Score, which will be used to compare the baseline model and the proposed model. The architecture can be seen in Fig. 3.
- The ResNet 50 backbone functions as the network's foundation. Multiple residual blocks are layered to

create a deep convolutional neural network. Each residual block is comprised of multiple convolutional layers, enabling the network to learn increasingly complex characteristics.

The skip connections in the residual blocks permit gradient flow during training, thereby resolving the vanishing gradient problem and enhancing the network's capacity to acquire meaningful representations. The ResNet 50 backbone analyzes the input image and extracts multiple levels of hierarchical features, capturing both low-level details and high-level semantic information. The PSPNet (Pyramid Scene Parsing Network) module is incorporated into the architecture to collect contextual data at multiple scales.

The module operates on the ResNet 50 backbone's generated feature maps. It utilizes a pyramid pooling mechanism to combine spatially distinct features. Pyramid pooling is accomplished by dividing the feature maps into multiple regions of differing sizes and then performing pooling operations (such as average pooling) within each region. By combining features at various dimensions, the PSPNet module captures both local details and global context, giving the network a comprehensive understanding of the image.

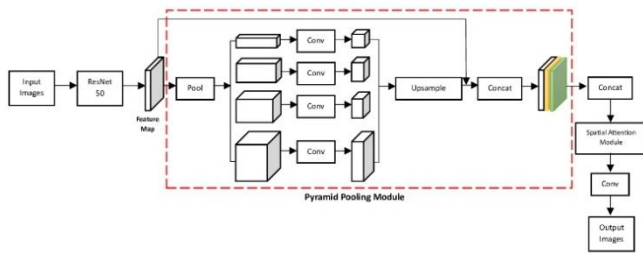


Fig. 3. The PRESSNet architecture.

The spatial attention mechanism is implemented to selectively highlight informative image regions. It improves the network's discriminative ability by focusing on pertinent features and suppressing irrelevant or chaotic regions. Indicating the significance of each location, the mechanism allocates attention weights to various spatial locations in the feature maps.

During the training process, these attention weights are learned, allowing the network to automatically attend to informative regions. The spatial attention mechanism helps the network make more accurate predictions and improves its overall performance by focusing on pertinent features. The ResNet-50 backbone in the PRESSNet architecture extracts complex image features. The PSPNet module then processes these features to capture contextual information at multiple scales. Lastly, the spatial attention mechanism selectively accentuates significant regions, thereby enhancing the network's capacity to concentrate on pertinent features.

By integrating these elements, the architecture is able to effectively capture both local and global context, extract high-level characteristics, and selectively focus on informative regions. This improves the performance of various image comprehension tasks, such as semantic segmentation, where precise object boundary delineation and accurate pixel-level predictions are essential.

#### IV. RESULT AND DISCUSSION

PRESSNet made use of the Palu earthquake disaster dataset, which it acquired from the xBD dataset. The proposed method, which is called "PRESSNet," combines the effectiveness of the PSPNet model with the architecture of "ResNet-50" and adds a "spatial attention" mechanism to get great results in tasks that have to do with understanding pictures. The ResNet-50 architecture is part of the PRESSNet framework. It uses deep residual blocks to collect high-level features and get discriminative representations in a way that uses few resources. The PSPNet model improves network performance by incorporating a pyramid pooling module that effectively combines contextual information from several dimensions. This enables the model to capture both local and global context, leading to enhanced performance. Furthermore, the integration of a spatial attention mechanism into PRESSNet enables the network to prioritize informative regions within the input image. This capability allows the network to concentrate on pertinent characteristics while effectively reducing the impact of noise and distractions. Through the integration of several components, the suggested PRESS-Net exhibits remarkable performance in the field of image analysis, showcasing the efficacy of including the PSPNet model alongside ResNet-50 and spatial attention mechanisms for the purpose of enhancing image processing tasks. The test results are displayed in Table I.

The assessment of various damage classes was conducted using three models: the baseline model (PSPNet+ResNet50), PRESSNet, and the PSPNet + ResNet 101 + Spatial Attention model. The evaluation revealed that the baseline model achieved an F1 score of 98.6% for the "Background" class. PRESSNet demonstrated a little superior performance compared to the base model, as seen by its F1 score of 98.62%. The F1 score achieved by the enhanced model PSPNet+ResNet101+spatial attention was 94.3%. The baseline model attained an F1 score of 88.6% for the "No Damage" category. The performance of PRESSNet was somewhat inferior to that of F1, as evidenced by its score of 88.2%.

The combination of PSPNet, ResNet 101, and spatial attention achieved a notable F1 score of 75.1%. The baseline model obtained an F1 score of 81.2% in the "Destroyed" class. The performance of PRESSNet showed a slight superiority compared to F1, as substantiated by its score of 82.4%. The PSPNet+ResNet101+spatial attention model had a significantly diminished performance, as evidenced by an F1 score of 28.3%. The observed substantial decrease in the F1 score pertaining to the "Destroyed" class, resulting from the utilization of an alternative backbone (PSPNet+ResNet101+spatial attention) in comparison to the baseline model (PSPNet+ResNet50) and PRESSNet, raises the inquiry regarding the underlying reasons for the drop in performance.

Several factors may have contributed to this performance loss that the PSPNet+ResNet101+Spatial Attention model is more complicated than the base model. It includes a more advanced ResNet architecture (ResNet 101) and spatial attention. It may have been more difficult to learn effective

representations for the “Destroyed” class due to the increased model complexity, resulting in a lower F1 score.

If the “Destroyed” class is not adequately represented in the training data, the model may struggle to acquire meaningful patterns for this class. Insufficient data can result in inaccurate generalizations and diminished performance, especially when dealing with uncommon or underrepresented groups.

**Imbalance in Class Distribution:** If the dataset is imbalanced, i.e., there is a significant difference in the number of samples between classes, the model’s performance may be negatively impacted. If the “Destroyed” class is underrepresented relative to other classes, the model may not have had sufficient exposure to acquire its unique characteristics, resulting in a lower F1 score.

To resolve the low F1 score for the “Destroyed” class in the improved model, it may be advantageous to investigate and analyze the specific difficulties and constraints posed by this class in greater depth. Obtaining more representative training data, meticulously balancing the class distribution, and refining the model architecture and hyperparameters could be potential solutions for improving the model’s performance on the “Destroyed” class.

The baseline model (PSPNet+ResNet50) received an F1 score of 89.5% based on the macro-average F1 scores, which provide a comprehensive performance measurement. With an F1 score of 89.7%, PRESSNet outperformed the baseline model by a small margin. The performance of the improved model, PSPNet+ResNet101+spatial attention, was 66% on the F1 scale.

Fig. 4 shows the result between the baseline model, PRESS-Net, and PSPNet + ResNet 101 + Spatial Attention. In general, PRESSNet performs comparably or slightly better than the baseline model (PSPNet+ResNet50) across all damage classes. However, the model that used ResNet 101 backbone did not improve performance across all classes, with the "Destroyed" class experiencing a significant performance decline.

It is essential to observe that these results are dependent on the evaluation metrics and data set employed. To comprehend the factors contributing to the performance disparities between models, additional analysis and investigation are required.

TABLE I. THE TEST RESULT

Class	Deep Learning Model		
	Baseline Model (PSPNet + ResNet 50)	PSPNet + Resnet 50 + Spatial Attention	PSPNet + Resnet 101 + Spatial Attention
F1 Background (Class 0)	0.986	0.986	0.943
F1 No Damage (Class 1)	0.886	0.882	0.751
F1 Destroyed (Class 2)	0.812	0.824	0.280
Macro Average F1	0.895	0.900	0.659

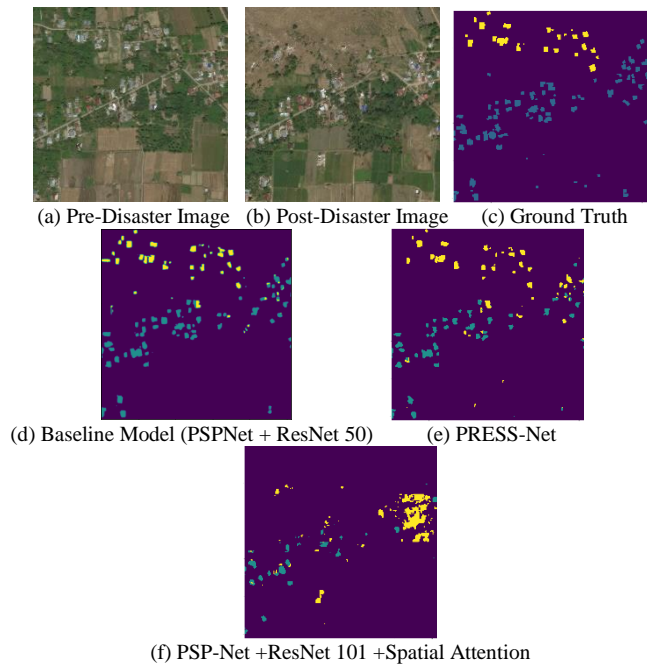


Fig. 4. The test result.

## V. CONCLUSION

This paper presents the PRESS-Net methodology for developing semantic segmentation models, with a particular focus on leveraging the Palu earthquake disaster dataset obtained from the xBD dataset. The PRESS-Net model achieved remarkable results in the field of image understanding by combining the beneficial aspects of the PSPNet model with the ResNet-50 architecture while also incorporating a spatial attention mechanism. The examined technique has demonstrated enhanced performance in comparison to the baseline model (PSPNet+ResNet50), particularly in its capacity to capture both local and global context information.

Upon conducting a comparative analysis of different versions of the model, this research has come across surprising findings. The baseline model (PSPNet+ResNet50) demonstrated commendable performance, achieving an F1 Score of 89.5%. The percentage was elevated to 89.7% as a result of the intervention of PRESS-Net. The F1 Score of the more intricate model, which combines PSPNet, ResNet101, and spatial attention, experienced a decrease to 66.0%. This decrease was particularly notable in the "Destroyed" category. The potential cause for this reduction could be attributed to the model's excessive complexity or the inadequate availability of suitable instances for learning. In order to tackle this matter, the research proposes the acquisition of more photos that portray scenes of destruction, ensuring a balanced distribution of image categories, and making modifications to the structure of the model.

In the future, there are several exciting research possibilities for improving building damage assessment. One option is to explore different types of deep learning models that might be better at recognizing structural damage. This could involve trying out various model designs or new methods in deep learning. Using larger and more diverse datasets can also

help the model work better in different situations. These datasets should cover a wide range of disasters and locations. To truly understand how well the model works in the real world, we need to test it with actual aerial and satellite images. Additionally, ongoing efforts to improve the model, like transfer learning or fine-tuning, can make a big difference when working with larger datasets. We should also consider adding more types of data, like weather or location information, to make the predictions more accurate. Lastly, we should validate the model's results in real-life situations to ensure it works effectively in disaster management. These suggestions are intended to guide future research in making deep learning models for building damage assessment more effective and reliable.

#### REFERENCES

- [1] Y. Wang, S. Ruan, T. Wang, and M. Qiao, "Rapid estimation of an earthquake impact area using a spatial logistic growth model based on social media data," *International Journal of Digital Earth*, vol. 12, pp. 1-20, 2018. doi:10.1080/17538947.2018.1549108.
- [2] Y. Du, L. Gong, Q. Li, and F. Wu, "Earthquake-induced building damage assessment on SAR multi-texture feature fusion," in *Proceedings of the IGARSS 2020—2020 IEEE International Geoscience and Remote Sensing Symposium*, pp. 6608-6610, 2020. doi:10.1109/IGARSS39084.2020.9323276.
- [3] C. Lin, Y. Li, Y. Liu, X. Wang, and S. Geng, "Building damage assessment from post-hurricane imageries using unsupervised domain adaptation with enhanced feature discrimination," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1-10, 2021. doi:10.1109/TGRS.2021.3060460.
- [4] R. Zhang, H. Li, K. Duan, S. You, K. Liu, F. Wang, and Y. Hu, "Automatic detection of earthquake-damaged buildings by integrating UAV oblique photography and infrared thermal imaging," *Remote Sensing*, vol. 12, pp. 2621, 2020. doi:10.3390/rs12162621.
- [5] S. Koshimura, L. Moya, E. Mas, and Y. Bai, "Tsunami damage detection with remote sensing: A review," *Geosciences*, vol. 10, pp. 177, 2020. doi:10.3390/geosciences10050177.
- [6] F. Nex, D. Duarte, F. G. Tonolo, and N. Kerle, "Structural building damage detection with deep learning: Assessment of a state-of-the-art CNN in operational conditions," *Remote Sensing*, vol. 11, pp. 2765, 2019. doi:10.3390/rs1123276.
- [7] Y. Endo, B. Adriano, E. Mas, and S. Koshimura, "New insights into multiclass damage classification of tsunami-induced building damage from SAR images," *Remote Sensing*, vol. 10, pp. 2059, 2018. doi:10.3390/rs10122059.
- [8] S. Garg et al., "Building damage assessment using deep learning from satellite imagery," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 133-140, 2017.
- [9] H. Hu et al., "Building damage assessment from aerial imagery using Siamese convolutional neural network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, pp. 3836-3849, 2018.
- [10] S. K. Shah et al., "Building damage assessment from satellite imagery using deep learning," *Remote Sensing*, vol. 11, pp. 2953, 2019.
- [11] K. He et al., "Deep residual learning for image recognition," *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016.
- [12] M. Waseem et al., "Building damage assessment using a hybrid deep learning approach from post-disaster satellite imagery," *Remote Sensing*, vol. 12, pp. 3674, 2020.
- [13] H. Zhao et al., "Pyramid scene parsing network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2881-2890, 2017.
- [14] Z. M. Hamdi, M. Brandmeier, and C. Straub, "Forest damage assessment using deep learning on high-resolution remote sensing data," *Remote Sensing*, vol. 11, pp. 1976, 2019.
- [15] W. Yuan, J. Wang, and W. Xu, "Shift pooling PSPNet: rethinking PSPNet for building extraction in remote sensing images from entire local feature pooling," *Remote Sensing*, vol. 14, pp. 4889, 2022.
- [16] J. Chen et al., "DASNet: Dual Attentive Fully Convolutional Siamese Networks for Change Detection in High-Resolution Satellite Images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 1194-1206, 2021.
- [17] V. Mnih et al., "Recurrent models of visual attention," in *NIPS*.
- [18] L. Chen et al., "Attention to scale: Scale-aware semantic image segmentation," in *CVPR*.
- [19] E. Weber and H. Kané, "Building disaster damage assessment in satellite imagery with multi-temporal fusion," 2020. doi:10.48550/arxiv.2004.05525.
- [20] A. Vaswani et al., "Attention is all you need," in *NIPS*.
- [21] X. Wang et al., "Non-local neural networks," in *CVPR*.
- [22] H. Zhao et al., "Psanet: Point-wise spatial attention network for scene parsing," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 267-283, 2018.
- [23] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 267-283, 2018.

# Group Intelligence Recommendation System based on Knowledge Graph and Fusion Recommendation Model

Chengning Huang<sup>1\*</sup>, Bo Jing<sup>2</sup>, Lili Jiang<sup>3</sup>, Yuquan Zhu<sup>4</sup>

School of Computer and Communication Engineering, Nanjing Tech University Pujiang Institute, Nanjing, 211222, China<sup>1,3</sup>  
School of Computer Science, Nanjing Audit University, Nanjing, 211815, China<sup>2</sup>  
School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang, 212013, China<sup>4</sup>

**Abstract**—The challenge of how to further improve the accuracy of the system's recommendations in a data-limited environment is crucial as the use of group intelligence recommendation systems in everyday life increases. Through the fusion of different types of auxiliary information, this study develops a multi-feature fusion model based on the conventional recommendation model by introducing knowledge graphs. It also considers the homogeneity of push results caused by graph convolutional network smoothing when using knowledge graphs, and designs a fusion label propagation algorithm and graph convolution. The multi-feature fusion model had a maximum hit rate of over 80% and a normalised discount gain of up to 43% running time much lower than the conventional graph convolution recommendation model in the representation dimension interval [2, 32], while the fusion label propagation algorithm and graph convolution network model maintained a hit rate and normalised discount gain higher than the conventional model by 2 to 1 under 10 consecutive epochs. With a hit rate and normalised discount gain 2 to 10 percentage points higher than the conventional model, the coverage rate increased to 49.8%. This study is useful for research on group intelligence recommendation systems and can serve as a technical guide for improving the ability of group intelligence systems to make recommendations quickly.

**Keywords**—Knowledge graphs; recommendation system; graph convolutional networks; label propagation algorithms

## I. INTRODUCTION

Group Intelligence Recommendation (GIR) System research is expanding along with the field of group intelligence technology. The common GIR systems predict the user's past choice data using neural network-like algorithms to create suggestions [1]. There is a pressing need to lessen the reliance of the recommendation model (RM) on users' previous data as standard recommendation systems have a tendency to produce highly biased outcomes when data information is limited [2]. Since in the actual recommendation process, in addition to information about the interaction between users and items, there is also information about user profiles, items, some relevant environmental conditions, etc., knowledge graphs (KGs) containing a wide variety of information data have been noticed. Multiple pieces of information can be combined thanks to the properties of KGs, which also improve the scalability of the recommendation system (RS) [3]. This is because, as a directed heterogeneous

graph, KG uses edges to represent relationships between entities and nodes to represent entities that can represent both users and items. As a result, interactions between users and other users as well as interactions between users and items can be fused into the graph as auxiliary information, which can be used to make up for the information deficit brought on by a data-scarce environment [4]. In general, conventional suggestion models possess certain limitations whereby they may encounter the issue of data sparsity. This refers to a scenario in which there is minimal interaction data between users and items. As a result, the accurate modelling of the relationship between users and items is challenging and may impede the correctness and customisation of suggestion outcomes. When the recommendation system begins operating, it lacks adequate user behavior data or item attribute information, over-relies on users' historical behavior data, and disregards their present interests and requirements. Therefore, it struggles to offer dependable personalized recommendations for new users or items. To design a multi-feature fusion model based on the traditional RM, this study attempts to improve the traditional RM. It also attempts to introduce KG while taking into account the problems with graph convolutional smoothing and homogenization of recommendation results that are easily encountered when applying KG. Finally, it attempts to introduce Label propagation algorithms (LPA) to Graph convolutional networks (GCN) in order to further optimise the multi-feature model. This research content has the potential to advance the field of recommendation systems and enhance the quality of recommendation services for practical applications.

The study is broken up into six sections: The second section gives a summary of the most recent research findings; the third section describes the study's methodology and design elements; the fourth section presents the experimental findings and an analysis based on those findings; and the fifth section summarises the study's findings and the prospect is given in sixth section.

## II. RELATED WORKS

People use GIR systems frequently, and with the growth of the web sector, customised RS has become increasingly important. According to Zhan et al, current recommendation models (RMs) primarily consider item compatibility modelling and do not consider user profiles, resulting in a

system where knowledge is pushed in a one-dimensional way without linking to user preferences. By adding attention to attribute-aware KG, Zhan et al. subsequently created an association between users and things, and created user-relationship-aware attention layers and goal-aware attention layers to capture user preferences. The results demonstrate the superiority of the model over other models for capturing user preferences [5]. Although Chen et al. argue that the current usage of GCNRM typically involves recursive aggregation with neighbouring nodes and their subsets, there is uncertainty in determining whether said neighbours can provide vital information after graph convolution. The introduction of KG in GCNRM is indeed beneficial in handling diverse multi-information tasks. Chen et al. proposed the Neighbour Enhanced Graph Convolutional Network (NEGCN) to enhance graph refinement process based on GCNRM and designed the neighbour evaluation method for critical information assessment. NEGCN demonstrated significantly improved model performance compared to the traditional graph convolutional RM [6]. Jiang et al. claim that the current RS approach to information exploitation is still limited and often only considers neighbourhood-specific information. To improve the conventional RS recommendation model, Jiang et al. propose a social aggregation neural network model based on attention mechanisms (AM). The model enables optimal user model embedding by propagating global social influence and capturing heterogeneous influences through AM. According to the results, using multi-layered perception to simulate the interaction between users and things is more flexible and successful than using conventional linear interaction algorithms to produce recommendation results [7].

Sang proposes a new knowledge graph-enhanced neural collaborative RM that can operate on information aggregation from multi-hop neighbours, and can use AM to grade the importance of relationships, as well as to model users and items in the embedding by modelling them in the embedding dimensional connections. On the other hand, the use of KGs in RM has always been hampered by the difficulty of modelling higher-order connectivity in large KGs using traditional models. The results showed that the model somewhat reduced the challenge of applying KG to RM [8]. Zhang et al. proposed a new knowledge-aware representation of the Graph Convolutional Recommendation Network model, arguing that in real recommendation environments, data is often sparsely distributed and the use of neighbourhood information alone is not sufficient to support accurate recommendation prediction results. The model can fuse item information through the propagation of links between nearby nodes in the KG and quickly capture correlations between people and items. This allows the model to predict likely user choices over time. This model, which creates user profiles by establishing neighbourhood associations between people and user objects by sorting features with high similarity between different users, has been shown to significantly outperform classical RM on a large dataset [9]. Graph neural networks, on the other hand, are favourable in dealing with complicated transitions between entity interactions in the environment of limited information input, according to Gwadabe et al. As a result, Gwadabe et al. proposed a graph neural network-based RM that may use graph neural networks to learn the ordered interactions first,

followed by the unordered interactions, in the current session. Numerous studies have revealed that this model performs noticeably better than other models in addressing unpredictable user behaviour in a data-limited environment [10].

In summary, experts have studied both the improvement of conventional RM and the implementation of KG, but have focused on optimising RM to increase the accuracy of push results to users, relying on the substantial information compensation found in KG itself. In reality, another problem with KG in RM is that a significant proportion of graph nodes are susceptible to convolutional smoothing, which can lead to homogenization of push results. It is still important to conduct research on how to use the entities' own information as much as possible in a data coefficient environment, and how to solve problems with the use of KG.

### III. DESIGN OF THE GIR MODEL FOR THE INTEGRATION OF KG

The principle of traditional RM is to get user behaviour prediction by analysing historical data of users, but the results derived from this model will be more deviant in a data sparse environment, so KG, which can fuse various kinds of auxiliary information, needs to be introduced to make up for the lack of interaction information. However, the KG itself is large in scale and is prone to the problem of excessive smoothing of a large number of graph nodes in the GCN, which can lead to homogenisation of the model's recommendations to users [11]. To address these problems, a multi-feature fusion model based on KG and AM and a model that fuses LPA and GCN are proposed.

#### A. Design of KGARA Model for Multi-Feature Fusion based on KG and AM

To address the limitations imposed on RM by data sparse environments, this research proposes the KGARA model. The core principle of this model is to incorporate relationship-aware structures to enrich the preference relationships between users and objects, and users and users. Based on joint AM, the model uses KG to fuse adjacent objects with different relationship types to obtain rich feature information. A graphical neural network is also used as a deep learning algorithm in RS.

The semantic information is first modelled using the representation-based KG recommendation algorithm, which vector embeds the input user features to obtain the initial representation, and then uses AM to complete the portrayal of user preferences. In this process, the representation is generated by the KG encoding operation on its entity, which requires the use of Knowledge Graph Embedding (KGE) [12]. KG, the interaction matrix of the relationship between the user and the item, and other information are fused to generate the user representation  $u$ , and the item representation  $v$ , and then they are inner-producted to obtain the probability of the user choosing the item, which is described by (1) is described.

$$\hat{y}_{uv} = u^T v \quad (1)$$

$T$  in (1) denotes the inner product operation and  $\hat{y}_{uv}$  is the user selection probability. Fig. 1 shows the structure of the model. From left to right, the first layer is an embedding vector layer, from which the unique hot codes of user, relationship and item are input and formed into an initial representation by vector embedding operation; then the unweighted KG formed by the initial representation is

transformed into a weighted KG by the attention layer and stored in the adjacency matrix; the next layer is the feature propagation layer, where the item representation is trained by GCN and fused with the domain representation using an aggregator; finally, the probability of the user selecting the item is obtained by inner product operation on the obtained item representation  $e_v$ .

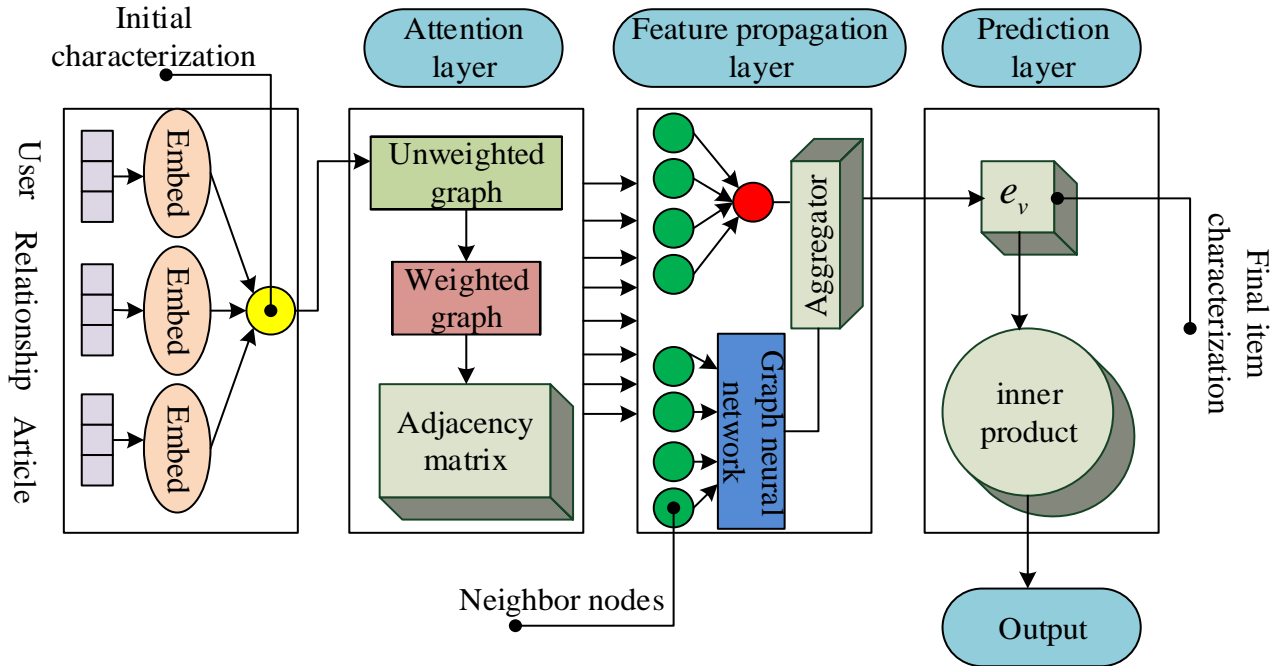


Fig. 1. Structure of the KGARA model.

In the relational attention network layer, the degree of importance between the relational representation and the user representation is obtained by doing an inner product operation on the two, described by (2).

$$c_u^r = e_u^T e_r \quad (2)$$

where  $e_r$  is the relational representation,  $e_u$  is the user representation and  $T$  denotes the inner product operation done on both. However, in the initial KG the edges can only describe the relationship but not the weight values, so the initial unweighted KG needs to be transformed into a weighted graph by (3).

$$y_{uv} = \begin{cases} 1, & \text{If } u, v \text{ interact;} \\ 0, & \text{If } u, v \text{ do not interact;} \end{cases} \quad (3)$$

$y_{uv}$  in (3) is a parameter in the user-item interaction matrix. The adjacency matrix resulting from the transformation into a weighted graph is denoted  $A_u$  and the relationship weights of entity  $i$  and entity  $j$  in row  $i$  and column  $j$  of this matrix are expressed in (4).

$$A_u^{i,j} = c_u^{r,i,j} \quad (4)$$

$c_u^{r,i,j}$  in (4) represents the entity relationships in the unweighted graph. Since the weighted graph itself can lead to an excessive computational burden, the KGARA model prioritises the nodes by attention weight values, which in turn yields the important nodes. The important nodes are weighted and summed by (5), which in turn gives the target entity representation.

$$e_{N(v)}^u = \sum_{\alpha \in N(v)} \hat{c}_\alpha^r e_\alpha \quad (5)$$

The set of target node  $v$  and neighbouring nodes in equation (5) is denoted by  $N(v)$ ,  $\alpha$  is a parameter taken from this set, and the normalised relational attention score is  $\hat{c}_\alpha^r$ . This indicates that nodes with a high relational attention score will be filtered out, as defined by (6) for them.

$$\hat{c}_\alpha^r = \frac{\exp(c_\alpha^r)}{\sum_{e \in N(v)} \exp(c_e^r)} \quad (6)$$

Next, at the feature propagation layer, the relational attention information is fused on the basis of the obtained weighted graph, and after all the rows in the matrix have been calculated, the individual entity representations are obtained,

denoted as  $h_k$ . The process is represented by equation (7).

$$\begin{cases} H_1 = \sigma(D_u^{-\frac{1}{2}} A_u D_u^{-\frac{1}{2}} H_0 W_0) \\ H_2 = \sigma(D_u^{-\frac{1}{2}} A_u D_u^{-\frac{1}{2}} H_1 W_1) \\ \dots \\ H_k = \sigma(D_u^{-\frac{1}{2}} A_u D_u^{-\frac{1}{2}} H_{k-1} W_{k-1}) \end{cases} \quad (7)$$

In (7), the representation matrix in row  $k$  is represented by  $A_u$ ,  $W_k$  is the parameter matrix in row  $k$ , and  $\sigma$  is the activation function. where there is a logarithmic matrix relationship between  $D_u^{-\frac{1}{2}}$  and  $A_u$ , described by (8).

$$D_u^{ij} = \sum_j A_u^{ij} \quad (8)$$

The item representation and the domain representation are aggregated by (8) to obtain the final item representation  $e_v$ . The final step of the model performs an inner product operation on the user representation and the item representation obtained from (8) to finally arrive at a probability value for that user to accept the recommended item, a process described by (9).

$$\hat{y}_{uv} = e_u^T e_v \quad (9)$$

### B. Design of A GCNLP Model Incorporating LPA and GCN

When applied to RM, the KG technique is computationally intensive due to the presence of tens of billions of edges and billions of nodes [13]. In GCNs, the large number of nodes can also cause the problem of smooth graph convolution and thus homogeneous recommendation results [14]. Therefore, in this study, the KGARA model is used to adjust the weight value of the edges of the GCN, and the attention network is used to filter the user's maximum weight on the items.

The structure of the GCNLP model is shown in Fig. 2. The structure of the feature propagation layer is improved from the structure of the KGARA model in Fig. 1, and the rest of the

embedding vector layer, attention layer, and prediction output layer are structured in the same way as the KGARA model. In the improved feature propagation layer, GCN operates on neighbouring nodes to derive item representations, while LPA is introduced to adjust the weight values for graph edges.

Specifically in the representation propagation layer, the model uses the GCN to make basic predictions and then uses the LPA to assist in adjusting the edge weights. As a multilayer feedforward neural network, the GCN is able to perform transformation and propagation operations on nodes in the graph, as described by (10).

$$P^{(k+1)} = \sigma(D^{-1} A P^k W^{(k)}) \quad (10)$$

In (9),  $\sigma$  is the activation function,  $\sigma$  is the parameter matrix of layer  $k$ , and the resulting  $P^{(k+1)}$  is the node representation of layer  $k$ . Present at nodes  $\sigma$  and  $v_j$ , the GCN to  $v_i$  update process is described by (11).

$$P_i^{(k+1)} = \sigma\left(\sum_{v_j \in N(v_i)} \alpha_{ij} p_j^{(k)} W^{(k)}\right) \quad (11)$$

In (11),  $P_j^{(k)}$  is the  $k$ th level node representation of the target node,  $\alpha_{ij}$  is the value of the  $i$ th row and  $j$ th column in the adjacency matrix, i.e. the weight value between nodes, and the  $i$ th node in the set of neighbours is represented by  $N(v_i)$ . After the domain nodes of the target node are aggregated to obtain the domain representation of the target node, they are then transformed into the next-order representation  $P_i^{(k+1)}$  of the target node, and (12) and Equation (13) provide a description of these two steps.

$$s_i^{(k)} = \sum_{v_j \in N(v_i)} \alpha_{ij} P_j^{(k)} \quad (12)$$

(13)  $\sigma$  is domain representation.

$$P_i^{(k+1)} = \sigma(s_i^{(k)} W^{(k)}) \quad (13)$$

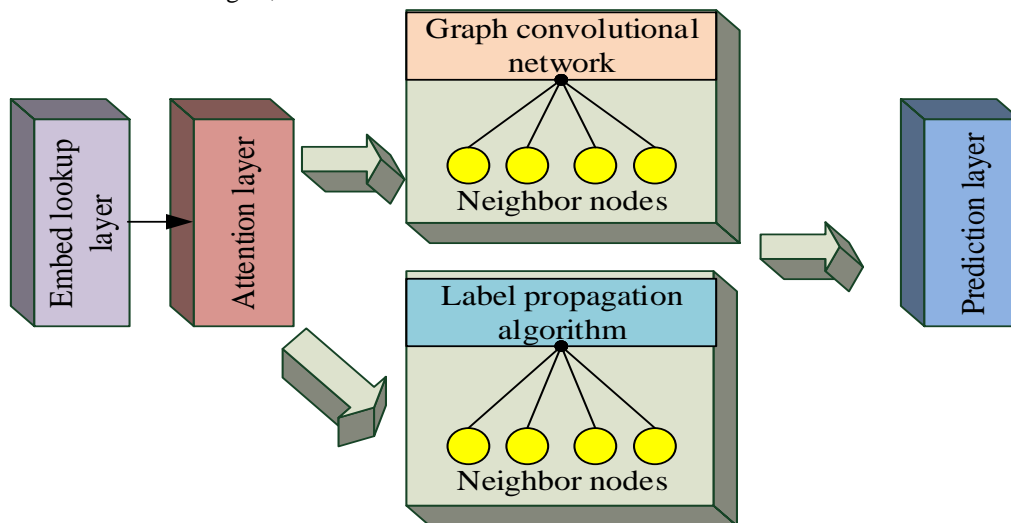


Fig. 2. GCNLP model structure.



The distance between  $v_i$  and  $v_j$  is defined as  $Q(p^{(k)})$ , and the distance between its domain representation  $s_i$  and  $s_j$  is defined as  $Q(s^{(k)})$ . After the aggregation operation on nodes  $v_i$  and  $v_j$ ,  $Q(s^{(k)})$  will be smaller than  $Q(p^{(k)})$  and similar nodes will be grouped into the same class, i.e. in the GCNLP model GCN places items of user attention in the same class to improve recommendation performance.

The propagation process of LPA at each level, i.e. according to the normalised edge weights, all nodes are subjected to label propagation by their neighbouring nodes and all nodes that already have labels are subjected to an initialisation operation by themselves to prevent the labels from disappearing [15]. The simulated label propagation process is shown in Fig. 3, assuming that the propagation process is performed only three times, with the goal of propagating from node a to node b. Red dots indicate with labels and colourless dots indicate without labels. In the first

execution (green line), node a passes the label to its neighbours node 3 and node 1, but node 1 already has a label, so the propagation path is inaccessible; in the second execution, node 3, which has already been propagated with a label, passes the label to its neighbours node b and node 4 (yellow line) and the first execution is completed, so node a initialises and propagates the label again (purple line); in the third execution, the nodes that already had labels in the previous execution also perform the initialization operation and continue to propagate the labels (blue line). Finally, the LPA must find all paths from node a to node b that are no longer than 3, as expressed by (14).

$$x_i^\infty = \sum_{j \in N(i)} a_{ij} x_j^\infty \quad (14)$$

In (13),  $x_i^\infty$  denotes the node in the label matrix,  $x_j^\infty$  is the value of column  $j$  of row  $i$  in the adjacency matrix, and  $N(i)$  is the set of neighbours of the node.

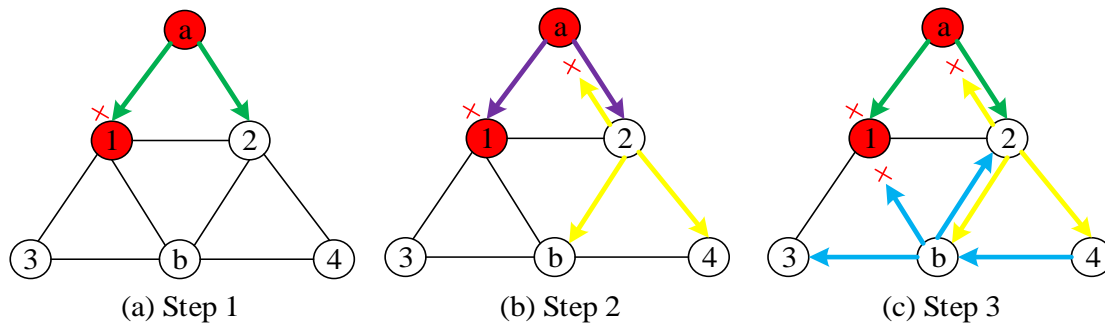


Fig. 3. Label propagation process.

The evaluation indicators introduced in this study were Hit Rate (HR), Normalized Discounted Cumulative Gain (NDCG) and Coverage, described by (15), (16) and (17) [16].

$$HR @ N = \frac{\text{Number of Hits @ } N}{|GT|} \quad (15)$$

*Number of Hits @ N* in equation (15) is the number of positive samples in the item sorting list in the recommendation task, and  $|GT|$  is the number of total samples in the test set.

$$NDCG @ N = \frac{\sum_{u \in U} NDCG_u @ N}{|U|} \quad (16)$$

Where  $|U|$  represents the number of users and  $\sum_{u \in U} NDCG_u @ N$  is the process of accumulating the normalised discounted gain in the test, resulting in a mean value of  $NDCG @ N$  [17].

$$\text{Coverage} @ N = \frac{|U_{u \in U} R(u)|}{|I|} \quad (17)$$

In equation (17)  $|I|$  is the set of items,  $U$  is the set of users, and  $RS$  is the list of items recommended by users

denoted by  $R(u)$ .

#### IV. EXPERIMENTAL RESULTS AND ANALYSIS

Performance testing experiments on the proposed KGARA model and the GCNLP model are conducted in the Book-Crossing environment, a book dataset, Movielens-1M, a film dataset, and Last FM, a music dataset [18]. For the model, the higher the value of HR and NDCG, the higher the quality. The Generalised Matrix Factorisation (GMF), Neural Matrix Factorisation (NeuMF) and Long Short-Term Memory R-GCN (LRGCN) were selected for the KGARA performance detection experiments [19]; the LRGCN, Ripple Net and Neural Graph Collaborative Filtering (NGCF) were selected for the GCNLP performance testing experiments [20]. An early termination strategy is implemented if HR@20 and NDCG@20 do not increase for 20 consecutive epochs on the test set, where an epoch is the process of training once using all samples in the training set.

##### A. Experimental Results and Analysis of the KGARA Model

First, parametric experiments were conducted to explore the effect of different representational dimensions on the model, followed by a comparison of the effect of the data sparse environment on the model performance. The parameter settings for the datasets in the experiments are given in Table I.

TABLE I. DATASET PARAMETER SETTINGS

Aggregator	BI-Interaction			
Data set	Learning rate	Batch size	Number of neighbors	Jump count
Configured parameters				
Book-Crossing	0.0002	32	8	1
Movielens-1M	0.02	2048	4	2
LastFM	0.0004	128	8	1

Fig. 4 shows the hit rate variation of the GMF, NeuMF, LRGCN and KGARA models on the three datasets under the representation dimension interval [2,32]. As can be seen from the figure, the overall hit rate of each model tends to increase as the representation dimension increases. However, compared to the Book-Crossing dataset and the Movielens-1M dataset, the increase for all models on the LastFM dataset varies between 1 and 3 percentage points, due to the fact that this dataset is less informative and can perform almost with sufficient information at lower representation dimensions. And in each dataset, compared to other models, the hit rate of KGARA proposed in this study is on average the highest, up to over 80% in the Movielens-1M dataset, which is on average 5 to 10 percentage points higher than the traditional graph neural network-based GMF and NeuMF; but in the Book-Crossing dataset, the average hit rate of LRGCN is

slightly higher than that of KGARA, which is also due to the simpler structural design of LRGCN than KGARA.

Fig. 5 shows the changes in the normalised discount gain of the GMF, NeuMF, LRGCN and KGARA models as the representation dimension increases on the three datasets. As can be seen in Fig. 5, the NDCG values of all models increase as the representation dimension increases, and although the increase in the NDCG metric is smaller, KGARA performs best when comparing both the initial NDCG values and the NDCG values at representation dimension 32: the initial NDCG value on the Movielens-1M dataset is around 42%, and as the representation dimension increases to 32, its NDCG value reaches around 43%. The average NDCG values of the KGARA model are much higher than those of the traditional graph convolution models GMF and NeuMF, ranging from 4 to 10 percentage points higher on average.

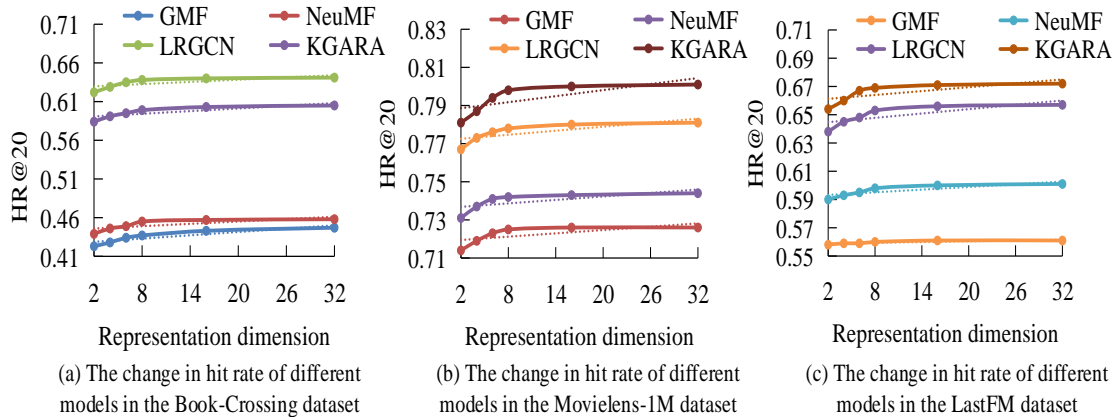


Fig. 4. Changes in hit rates of various models on different datasets.

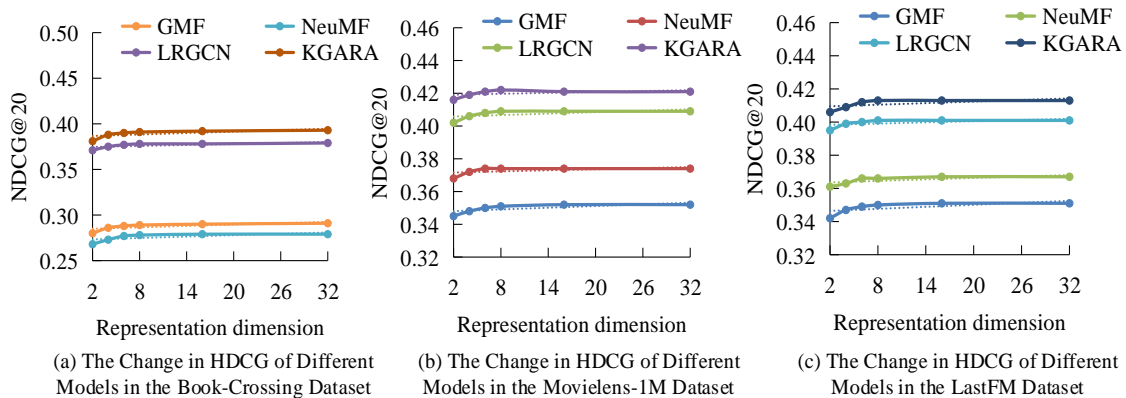


Fig. 5. Normalized loss gain changes of each model on different datasets.

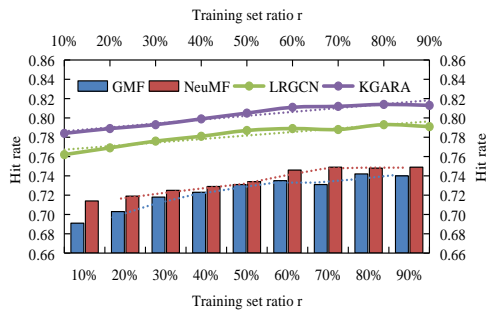


Fig. 6. Changes in hit rates of various models in data scarcity environments.

To test the performance of the KGARA model with the introduction of the KG feature fusion structure in a data sparse environment, the GMF, NeuMF, LRGCN and KGARA models were tested on the Movielens-1M dataset, and the hit rate variation of all models was compared by adjusting the proportion of the training set to the test set for this dataset. The experimental results are shown in Fig. 6, where  $r$  represents the proportion of the training set. As can be seen from Fig. 5, the hit rate of KGARA reached more than 78% when the proportion of the training set was only 10%, which was 2 to 10 percentage points higher than the other models, and as the proportion of the training set increased, the hit rate of KGARA also increased, stabilizing at 81.6% when the proportion of the training set reached 60%, with the highest hit rate reaching 81.8%. The other models basically stabilized after the training set percentage increased to 70%, and the final hit rate was still lower than KGARA by 2 to 8 percentage points.

Next, the runtimes of all models were examined on the

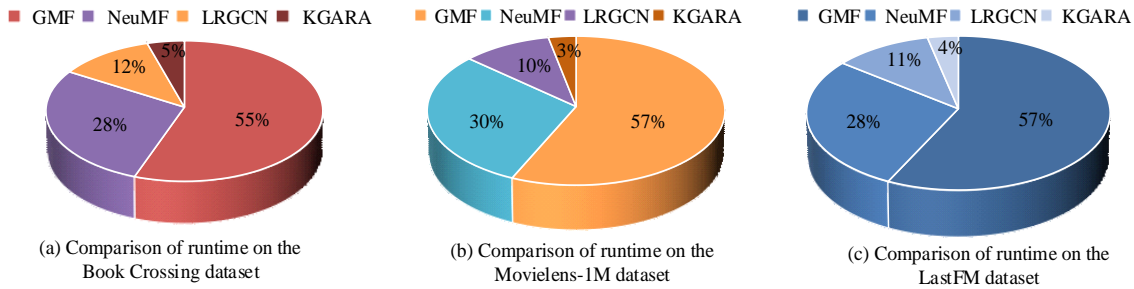


Fig. 7. The proportion of running time of each model on different datasets.

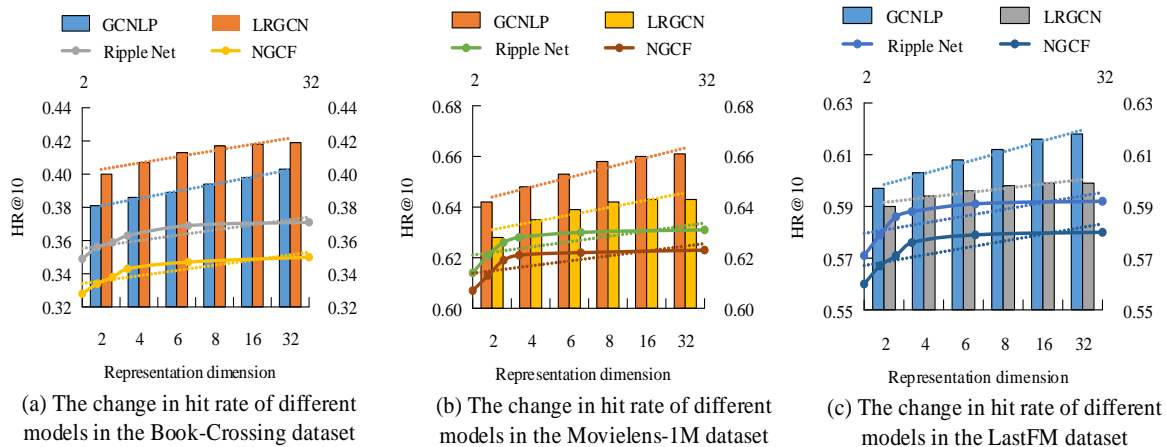


Fig. 8. Changes in hit rates of various models on different datasets.

three datasets at 1600MHz core frequency, RTX2080 graphics card and 8G video memory. Fig. 7 shows the results of the runtime comparison for all models. As can be seen from Fig. 7, although the runtime ratios of the models varied from dataset to dataset, the largest average runtime ratio was for the traditional graph convolution model GMF, which accounted for 56% of the four models; the model with the lowest runtime ratio was KGARA, which accounted for 4% on average, almost a tenth of the GMF runtime.

### B. Experimental Results and Analysis of the GCNLP Model Doing the Results Analysis

In addition to testing the basic performance of GCNLP in different data dimensions, this study also needs to test whether GCNLP can improve the problem that GCN is prone to homogeneous recommendation results due to graph convolution smoothing, so 10 consecutive epochs were analyzed for the GCNLP model in terms of three metrics: hit rate, normalized discount gain and coverage.

Fig. 8 shows the hit rate variation of the four models GCNLP, LRGCN, Ripple Net and NGCF on different datasets in the environment with representation dimensions {2,4,6,8,16,32}. As can be seen from the figure, although the hit rate on the Book-Crossing dataset is on average 2 percentage points higher than that of the GCNLP model due to the simple structure of LRGCN, the GCNLP hit rate is 2 to 4 percentage points higher than the other models on all other datasets, with the GCNLP hit rate on the Movielens-1M dataset being on average the highest, averaging in the upper 65% range with a maximum of 66%.

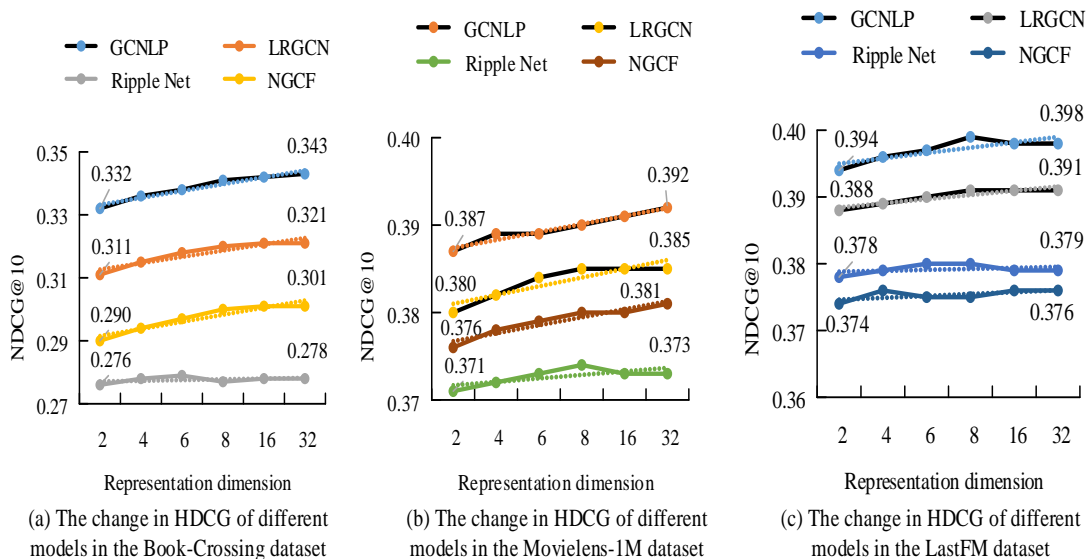


Fig. 9. NDCG changes of each model on different datasets.

Fig. 9 shows the variation in normalised discount gain for each model on the different datasets, and it can be seen that the NDCG values for all four models increase as the characterisation dimension increases on all three datasets, but the NDCG values for GCNLP are higher than the initial and final NDCG values for the other models compared to the other models. GCNLP had the highest NDCG values overall on the LastFM dataset, averaging around 36.5% and up to 40% over the period, while being one to two percentage points higher than the other models.

For GCNLP improvements made on the basis of the KGARA model, the metric coverage (Coverage) can be targeted to detect GCNLP performance. The higher the value of  $Coverage@N$ , the higher the coverage of the model and the less likely the problem of homogeneous recommendation results will occur. Fig. 10 provides a comparison between the mean coverage of all models under the low representation dimension and the mean coverage under the high

representation dimension based on three distinct datasets. As observed in Fig. 10, the coverage of each model increases as the representation dimension increases. The GCNLP and LRGCN show comparable coverage in the low representation dimension. However, in the high representation dimension, GCNLP exhibits a higher average coverage than all other models.

Table II displays comprehensive average coverage information for each model in both high and low representation dimensions. It is evident that, in the high representation dimension, the Book-Crossing dataset showed the highest coverage of GCNLP at 49.8%, while, in the low representation dimension, the Movielens-1M dataset presented the highest coverage of GCNLP at 41.2%. Overall, GCNLP's coverage was greater than the other models, indicating that the enhancements made to GCN using LPA improved the homogeneity of RM push results.

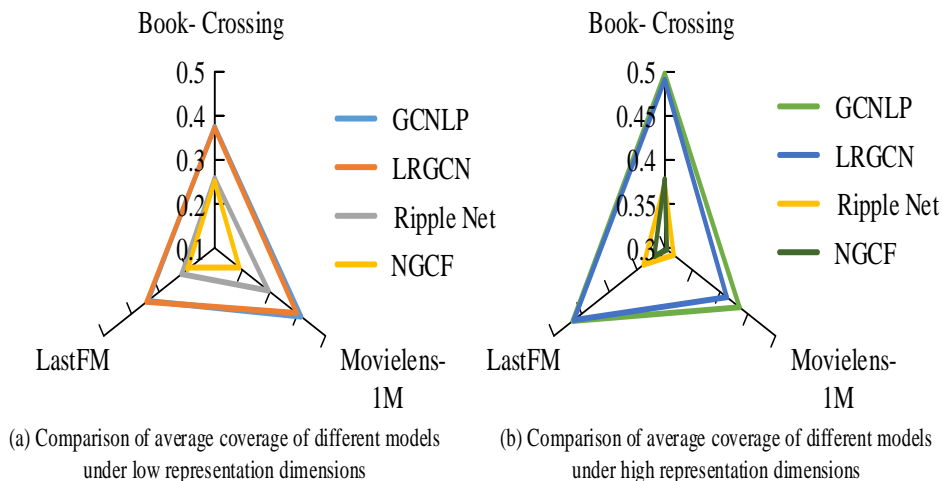


Fig. 10. Comparison of average coverage of various models.

TABLE II. DETAILED DATA ON AVERAGE COVERAGE OF EACH MODEL

Under low feature dimensions	Book-Crossing	Movielens-1M	LastFM
GCNLP	0.375	0.412	0.341
LRGCN	0.373	0.396	0.345
Ripple Net	0.258	0.294	0.219
NGCF	0.251	0.189	0.198
Under high feature dimensions	Book-Crossing	Movielens-1M	LastFM
GCNLP	0.498	0.435	0.466
LRGCN	0.491	0.412	0.464
Ripple Net	0.373	0.316	0.338
NGCF	0.378	0.304	0.319

## V. CONCLUSION

The aim of this study was to address the problem of large deviations in recommendation results in traditional GIR models within a data sparse environment. To achieve this goal, the study developed a multi-feature fusion model, KGARA, and improved its GCN structure by combining with LPA to obtain the GCNLP model. This approach enabled a more accurate and reliable model, thus mitigating the issue mentioned above. The study also introduced KG and AM based on traditional RM to compensate for the lack of interaction information. The results of 20 consecutive epoch trials show that the KGARA model has a maximum hit rate of over 80% and a normalised discount gain of over 43% when the characterisation dimension interval [2,32] is shifted, which is higher than other models. Furthermore, the proportion of the Movielens-1M training set was altered from 10% to 90%, with a maximum hit rate of 81.8%, ultimately illustrating the KGARA model's superior performance compared to all other comparative models in the data-scarce setting. Owing to its efficient nature and the shortest running duration among its counterparts, the KGARA model proves to be the most effective. Based on the results from 10 consecutive epochs of experimentation on the GCNLP model, the hit rate and normalised discount gain surpass other comparison models when the representation dimension is altered within the range of [2,32]. Specifically, the hit rate reaches up to 66% and the normalised discount gain reaches up to 40%, both of which are higher than those of other models. Moreover, the GCNLP model attained an average coverage of 41.2%, which outstripped the conventional graph convolution model's average coverage by a significant margin. This suggests that the addition of LPA to the GCN structure was a fruitful improvement and could potentially resolve the problem of homogenization in push outcomes. The study successfully enhanced the classic push model; however, it only considered user preferences' constant conditions and omitted dynamic shifts in user preferences. This aspect warrants future investigation.

## VI. PROSPECT

The KGARA and GCNLP models exhibit higher hit rates, normalized discount gains, and coverage performance when compared to other models. These outcomes suggest that the improved models are anticipated to outperform

recommendation systems, enhancing the standard and personalization of recommendation outcomes. Such accomplishments can offer motivation and establish benchmarks for the betterment and refinement of recommendation system models. One potential avenue for future research could involve the development of recommendation algorithms that rely on temporal data. Such algorithms could utilize users' past behavior and historical data to predict their future interests. This could involve the application of techniques including time series analysis, sequence modeling, and deep learning.

## ACKNOWLEDGMENT

The research is supported by: National Natural Science Foundation Program of China: Research on Distributed Progressive Classification Mining Method for Big Data Based on Reuse of Existing Knowledge(61702229); National Natural Science Foundation Program of Jiangsu: Key Technology Implementation of Smart Medical Platform in the Context of Big Data Cloud Computing(18KJD520001); Research Key Cultivation Project of Pujiang College, Nanjing University of Technology: Research on Multiple Personality Recommendation by Integrating Knowledge Graph and Attention Mechanism(njj2022-1-07); Nanjing University of Technology Pujiang College Youth Teacher Development Fund (PJYQ03).

## REFERENCES

- [1] B. Hu, Y. Ye, Y. Zhong, J. Pan, and M. Hu, "TransMKR: Translation-based knowledge graph enhanced multi-task point-of-interest recommendation," *Neurocomputing*, vol. 474, no. 14, pp. 107-114, 2022.
- [2] J. Chen, B. Li, J. Wang, Y. Zhao, L. Yao, and Y. Xiong, "Knowledge Graph Enhanced Third-Party Library Recommendation for Mobile Application Development," *IEEE Access*, vol. 8, pp. 42436-42444, 2020.
- [3] Z. Hu, J. Wang, Y. Yan, P. Zhao, J. Chen, and J. Huang, "Neural graph personalized ranking for Top-N Recommendation," *Knowledge-Based Systems*, vol. 213, no. 8, pp. 106426.1-106426.9, 2020.
- [4] Z.Y. Zhang, L. Zhang, D.Q. Yang, and L. Yang, "KRAN: Knowledge Refining Attention Network for Recommendation," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 16, no. 39, pp. 1-20, 2022.
- [5] H. Zhan, J. Lin, K.E. Ak, B. Shi, L. Duan, and A.C. Kot, "A3-FKG: Attentive Attribute-Aware Fashion Knowledge Graph for Outfit Preference Prediction," *IEEE Transactions on Multimedia*, vol. 24, pp. 819-831, 2022.

- [6] H. Chen, Z. Huang, Y. Xu, Z. Deng, F. Huang, P. He, and Z. Li, "Neighbor enhanced graph convolutional networks for node classification and recommendation," *Knowledge-based systems*, vol. 246, no. 21, pp. 108594.1-108594.101, 2022.
- [7] N. Jiang, L. Gao, F. Duan, J. Wen, T. Wan, and H.L. Chen, "SAN: Attention-based social aggregation neural networks for recommendation system," *International Journal of Intelligent Systems*, vol. 37, no. 6, pp. 3373-3393, 2021.
- [8] L. Sang, M. Xu, S. Qian, and X. Wu, "Knowledge graph enhanced neural collaborative recommendation," *Expert Systems with Applications*, vol. 164, no. 12, pp. 113992.1-113992.13, 2021.
- [9] L. Zhang, Z. Kang, X. Sun, H. Sun, B. Zhang, and D. Pu, "KCREC: Knowledge-aware representation Graph Convolutional Network for Recommendation," *Knowledge-Based Systems*, vol. 230, no. 27, pp. 107399.1-107399.13, 2021.
- [10] T.R. Gwadabe and Y. Liu, "Improving graph neural network for session-based recommendation system via non-sequential interactions," *Neurocomputing*, 2022, 468(Jan.11):111-122.
- [11] J. Chen, J. Yu, W. Lu, Y. Qian, and P. Li, "IR-Rec: An Interpretive Rules-guided Recommendation over Knowledge Graph," *Information Sciences*, vol. 563, no. 10, pp. 326-341, 2021.
- [12] Z. Yang and S. Dong, "HAGERec: Hierarchical Attention Graph Convolutional Network Incorporating Knowledge Graph for Explainable Recommendation," *Knowledge-Based Systems*, vol. 204, no. 27, pp. 106194.1-106194.11, 2020.
- [13] L. Zhang, Z. Kang, X. Sun, H. Sun, B. Zhang, and D. Pu, "KCREC: Knowledge-aware representation Graph Convolutional Network for Recommendation," *Knowledge-Based Systems*, vol. 230, no. 27, pp. 107399.1-107399.13, 2021.
- [14] H. Yang, H. He, W. Zhang, and Y. Bai, "MTGK: Multi-Source Cross-Network Node Classification via Transferable Graph Knowledge," *Information Sciences*, vol. 589, no. 1, pp. 395-415, 2022.
- [15] A. Salamat, X. Luo, and A. Jafari, "HeteroGraphRec: A heterogeneous graph-based neural networks for social recommendations," *Knowledge-Based Systems*, vol. 217, no. 4, pp. 106817.1-106817.10, 2021.
- [16] Y. Lin, B. Xu, J. Feng, H. Lin, and K. Xu, "Knowledge-enhanced recommendation using item embedding and path attention," *Knowledge-Based Systems*, 2021, 233(Dec.5):107484.1-107484.11.
- [17] S. Tao, R. Qiu, Y. Ping, and H. Ma, "Multi-modal Knowledge-aware Reinforcement Learning Network for explainable recommendation," *Knowledge-Based Systems*, vol. 227, no. 5, pp. 107217.1-107217.11, 2021.
- [18] S. Oslund, C. Washington, and A. So, et al., "Multiview Robust Adversarial Stickers for Arbitrary Objects in the Physical World," *Journal of Computational and Cognitive Engineering*, vol. 1, no. 4, pp.152-158, 2021.
- [19] N. Nassar, A. Jafar, and Y. Rahhal, "A novel deep multi-criteria collaborative filtering model for recommendation system," *Knowledge-Based Systems*, vol. 187, pp. 104811.1-104811.7, 2020.
- [20] S. Qianna, "Evaluation model of classroom teaching quality based on improved RVM algorithm and knowledge recommendation," *Journal of Intelligent and Fuzzy Systems*, vol. 40, no. 2, pp. 2457-2467, 2021.

# Statistical Language Model-based Analysis of English Corpora and Literature

Wenwen Chai\*

School of Foreign Languages, Zhengzhou Normal University,  
Zhengzhou, 450044, China

**Abstract**—Despite widespread use of statistical language models in language processing, their ability to process natural languages is not advanced and they struggle to effectively capture linguistic information. Furthermore, there is a lack of automatic processing models in the field of natural language processing. In order to address these issues, and improve the processing ability of statistical language models for English language a statistical language model optimization algorithm has been proposed. This algorithm is based on an improved resorting algorithm and is specifically applied to process English literary texts. Experimental results indicate that the proposed algorithm outperforms the N-gram algorithm in a majority of texts, with a maximum accuracy improvement of 14.5%. Additionally, in terms of the grammar analysis model, there is a high level of consistency between the model's scoring and the expert manpower scoring, as reflected by a correlation coefficient of 0.7893. This high level of consistency between the grammar analysis model and expert analysis results holds significant importance for the advancement of natural language processing.

**Keywords**—Statistical language model; corpus; English literature; reordering; grammatical analysis

## I. INTRODUCTION

Currently, utilizing automated algorithms to process natural language is one of the important research topics in the fields of corpus and translation. Statistical language models are models that use statistics to calculate the probability distribution of word occurrences in a particular language or context, which users use as a basis for operations and predictions [1]. With the maturity of technologies such as machine translation and speech recognition, statistical language models have become more widely used [2]. However, as a data-driven model, a single statistical language model has limited ability to process natural language and cannot reflect the linguistic features of natural language [3]. Based on the limitations of the statistical language model, various natural language processing algorithms that have applied the model also tend to be far less capable than human analysis [4]. In order to effectively improve the statistical language model's ability to process natural language and to apply it to natural language processing work, a reordering algorithm based on an improved minimum error training method is proposed. The reordering is an optimization technique, which can optimize the output of statistical language models by reordering the phrases [5]. A grammar analysis algorithm for English literature is proposed based on the reordering algorithm. Overall, this study proposes an English prediction and literary analysis algorithm based on statistical language models. This model aims to effectively

enhance the processing ability of statistical language models for English natural language.

This article is divided into seven sections. The second section introduces the research progress in related fields. The third section introduces the construction ideas and process of the model. The fourth section is the display of experimental results. The fifth section is the discussion. The sixth section is the conclusion. Lastly, seventh section discusses the limitations and future work.

## II. REVIEW OF THE LITERATURE

Statistical language models, one of the most important models in the field of natural language processing, have been studied and applied currently. Desai and his team examined the eye and neural activity of forty subjects during reading activities based on statistical language models combined with medical tests and found that the processing cost of low-frequency words was reduced due to contextual cues. The meanings of high-frequency words were more easily accessible and integrated with context [6]. The research results provide results based on human science for the processing of natural language. Teks P led his research team to conduct a machine translation study for Lampung Nyo dialect and compared the approaches based on statistical language models [7]. The project aimed to help student immigrants in Lampung province to translate the Lampung dialect of Nyo through the model and the proposed method was adopted as a working model with an accuracy rate of 59.85%. Sreelekha and Bhattacharyya [8] provided a solution for machine translation of Indian languages where digital resources are scarce by using Indowordnet lexical database to extend statistical language models and evaluate 440 models for 110 pairs of languages for comparison. They found that using lexical database mapping helped to resolve linguistic ambiguities and improve translation quality. Collins et al. [9] provided a framework for processing communication language data based on statistical language models using generalized linear mixed models and Bayesian methods, which, based on the results of the sample analysis, was able to analyze and compare the discourse patterns of children who had experienced traumatic brain injury and typically developing children differences between them. This study has important implications for the field of language processing and the study of childhood brain injury. Ycel et al. [10] used statistical language models to construct a computer-based system for learning foreign language vocabulary. They used specified software to display various card sets constructed using the proposed algorithm and examined the polysemantic correlations between behavioral variables and difficulty levels

of different word categories. This study provides an effective method for learning foreign language vocabulary. The author in [2] investigated the specific case of word frequency effects decreasing with age based on word frequency theory in statistical language models and suggested that word frequency effects may occur at different stages of language production. Ge [11] proposed a hybrid research framework combining word frequency analysis from Google Books Ngram Viewer with other analyses in conjunction with statistical language models, aimed at developing a linguistic and cultural concept analysis. Their findings showed a strong correlation between languages in different regions and their cultural concepts, and the frequency of concept words indicated a stronger collectivist culture in China compared to the U.S. Poncelas et al. [12] proposed a feature decay extension algorithm based on a parallel corpus and a statistical language model in order to delve into feature decay algorithm techniques to achieve a better method of training data instance selection. This method can reduce the execution time of FDA and improve the translation quality when multiple computational units are available. This study provides an important reference for improving the performance of machine translation using FDA technology.

A review of recent research related to statistical modeling of language reveals that most of the research in this field focuses on machine translation. In addition, some studies have combined statistical language models with the fields of medicine and sociology. In the field related to statistical language models, there are fewer studies investigating how to improve their ability to recognize natural language, and there is a lack of related applications in the last three years. Based on this gap area, this research focuses on the improvement of statistical language models and their application in the field of natural language recognition.

### III. ENGLISH CORPUS OPTIMIZATION AND LITERARY ANALYSIS BASED ON STATISTICAL LANGUAGE MODELS

#### A. A Statistical Language Model-Based Algorithm for Reordering English Corpus Output

Statistical language models calculate the frequency of occurrence of these concepts in a corpus based on the historical data of a given sequence of words and the likelihood of each word in that sequence. Although this technique is currently widely used in areas involving language processing such as speech recognition, and its translation, statistical language models, as a data-driven model, have biases in the estimation of real natural language [13]. This is due to the limitation of data size and data content. Lexical models, N-gram models, and co-occurrence models are all reordering models that have emerged to make statistical language models closer to real natural language [14]. However, the degree of fit of these models to natural language still needs to be optimized. In this study, a reordering method based on minimum error rate training is proposed. Minimum error rate training is a theory applied to the field of machine translation, but it can be improved and applied to this English corpus optimization and

literary analysis. In the English to other languages literary analysis scenario, the results of the statistical linguistic model-based translation for a specific utterance are shown in (1).

$$\hat{R} = \arg \max \Pr(R|f) \quad (1)$$

In (1)  $\hat{R}$  is the output result,  $f$  is the original utterance to be processed, and  $R$  is the output target language utterance. To obtain the output with the lowest error rate, the log-linear model is used to compute the posterior probability of the sentence pair  $(R, f)$  and recalculate the score, i.e., the ranking basis. The calculation procedure is shown in (2).

$$S(R, f) = \Lambda \bullet \Phi(R, f) \quad (2)$$

In (2),  $S(R, f)$  is the score,  $\Phi(R, f)$  is the feature vector linking the log-linear model and the sentence pairs, and  $\Lambda$  represents the weights of all features. Then the posterior probability can be defined as (3).

$$P(R|f) = \frac{\exp(S(R, f))}{\sum_{R'} \exp(S(R', f))} \quad (3)$$

Based on the results of the recalculated scores and the posterior probabilities, the system reorders the candidate results and outputs the new optimal items as shown in (4).

$$\hat{R} = \arg \max \Pr(R|f) = \arg \max S(R, f) \quad (4)$$

In the process of minimum error rate training, feature parameter weights need to be tuned and determined. The session first requires giving each parameter an initial value of weight and debugging for individual parameters. The other non-object parameters are treated as constants during debugging. Next proceed to apply the parameter in to other sentences of the corpus. The process is shown in Fig. 1. Fig. 1(a) shows the tuning process of one parameter  $a$  on the optimal solution selection of sentence R1. Different parameters take different value intervals corresponding to different optimal solutions. Fig. 1(b) depicts the test results of parameter  $a$  corresponding to sentence R1 in other sentences.

After completing this test, all segmentation points are identified and the optimal values of all sentences are found between each segmentation point. The next step is to perform error statistics for the optimal values in each interval, as shown in Fig. 2. Fig. 2 shows the total number of errors statistics for parameter  $a$ . As the value of  $a$  varies, the total number of errors statistics also fluctuates significantly, with smaller total number of errors representing better results from the statistical language model output. After following this process for all parameters, the whole algorithm is iterated until the error value statistics tend to be stable, which is more desirable.



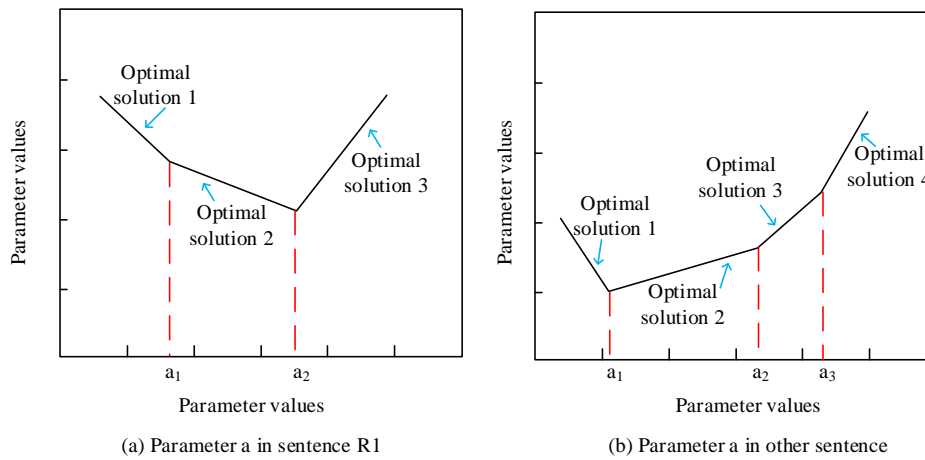


Fig. 1. Adjustment process of feature parameter weights.

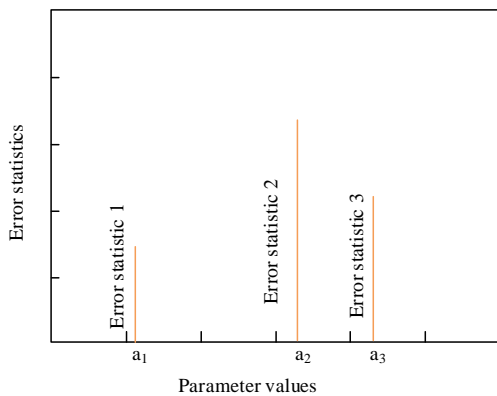


Fig. 2. Count of total error of different parameters.

In order to further enhance the performance of reordering and optimize the results, two sub-models with embedded minimum error rate training are proposed. The sub-models include lexical indication model and lexical N-element co-occurrence model. The lexical indication model performs lexical classification work for the statistical language model. Accurate lexical classification is the basis for the statistical language model to work properly and perform correct literary analysis. There are many possible lexical sequences for a word string, and some of the traditional models directly output the most common lexical properties of words. This method is the most cost-efficient and fast, but the accuracy rate is not satisfactory. To improve this situation, a lexical indication model is considered using a hidden Markov model. The hidden Markov model is shown in Fig. 3, which consists of hidden sequences, observed sequences and different probability distributions. According to the structure of this model, the selection of parameters directly affects the model performance. It has three main parameters, which are noted here as  $\lambda = (\pi, a, b)$ . The lexical indication task can be analogized to a decoding problem, i.e., finding the optimal sequence of states based on a given word sequence to generate a sequence of observations and a set of parameters. Hidden Markov models can efficiently solve such decoding problems.

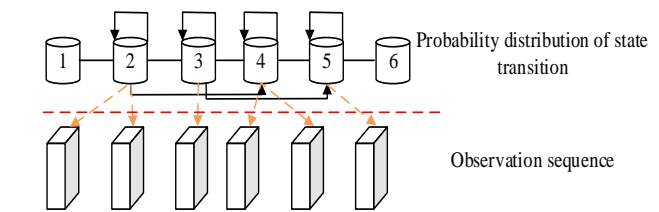


Fig. 3. Hidden Markov model.

The lexical N-element co-occurrence model is to integrate lexicality into the traditional word N-element co-occurrence model. The traditional word N meta model calculates the probability distribution by lexicon, while the lexical N meta co-occurrence model calculates it by lexicality, as shown in (5).

$$P(T) = \prod_{i=1}^n p(t_i | t_1, \dots, t_{i-1}) \quad (5)$$

In (5),  $p(t_i | t_1, \dots, t_{i-1})$  represents the lexical N\$ probability.  $t_i$  represents the different lexical properties. After obtaining the lexical N-probability, we need to deal with the co-occurrence relationship between different words. The co-occurrence is when two words appear together, and the more co-occurrence of two words in the text, the stronger the connection between them. In the lexical N meta co-occurrence model, instead of word-to-word co-occurrence, word-to-word co-occurrence is used, as shown in (6).

$$P(T|W) = \prod_{i=1}^n p(t_i | w_i) \quad (6)$$

In (6),  $W$  is a word sequence and  $T$  is its corresponding lexical sequence. Correspondingly, the co-occurrence frequencies of words and lexemes are shown in (7).

$$P(W|T) = \prod_{i=1}^n p(w_i | t_i) \quad (7)$$

The two sub-models are embedded in the minimum error training with linear interpolation, and the optimal results are re-output using linear re-ordering. Specifically, when the statistical language model based on minimum error training

outputs the ranking results, the two sub-models process the output word order with probability calculation, and then the probability calculation results are linearly interpolated with the ranking results of the statistical language model, as shown in (8).

$$P(W) = c_1 p_1(W) + c_2 p_2(2W) + c_3 p_3(W) \quad (8)$$

In (8),  $P(W)$  is the recalculated probability.  $c_i$  is the weight of the sub model, and  $P_i$  is its probability. This completes the construction of the proposed reordering algorithm, which utilizes two sub-models for optimization and is able to output results that are closer to natural language than the general reordering model.

### B. English Grammar Evaluation Model for Literary Analysis

The analysis of English literature has been one of the important application areas of statistical language models [15]. Due to the complexity and variability of natural language, algorithm-based literary analysis has been more difficult [16]. In this study, a grammar evaluation model based on statistical language models is proposed for the grammar evaluation aspect of English literary analysis. The model applies the proposed minimum error training reordering algorithm and incorporates the Transformer structure. The Transformer structure is an encoder-decoder model as shown in Fig. 4. The structure consists of six identical decoders with sub-layers. Each sublayer is connected with a normalization module and residuals between them [17]. There are two types of sub-layers, the fully connected network layer and the attention mechanism layer [18]. The number of layers of encoder and decoder is adjustable under this structure [19]. Considering the cost and

computational consumption, the number of layers of both encoder and decoder is set to 6 here.

In written English literature, most of its grammar is fluent and correct, and the problematic ones are usually small. Therefore, the Transformer model is used to move the sentences without grammatical problems directly to the target sentences, thus avoiding the interference of the grammar evaluation model with the sentences without grammatical problems. The mechanism of the probability distribution of words in the target sentence is shown in (9).

$$P_t(W) = a_t p_t^{copy}(w) + p_t^{gen}(w)(1 - a_t) \quad (9)$$

In (9),  $P_t(W)$  is the lexical probability distribution in the target sentence.  $P_t^{gen}$  is the probability distribution of grammar evaluation generation, and  $P_t^{copy}$  is the probability distribution of original utterance replication.  $a_t$  is the parameter used to control the probability of generation and replication at each time  $t$ . The Transformer structure is used in English grammar evaluation in the way shown in Fig. 5. The Transformer model itself is used to generate the probability distribution of the target vocabulary. The replication score is then calculated by the joint determination of the original utterance input this and the implicit state of the target word. The concept of attention mechanism of the Transformer model needs to be introduced here. The attention mechanism solves the problem of interaction, selection and integration between multiple information sources. It enables the model to focus more on the parts of high importance in the operation. Under the attention mechanism, sentences with a higher probability of grammatical problems are given higher weights.

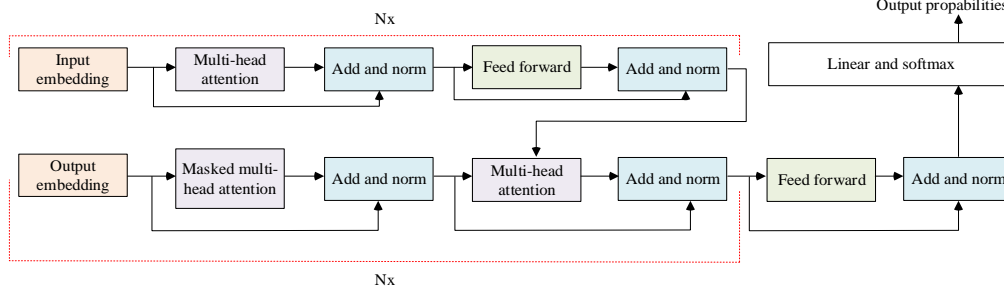


Fig. 4. Transformer structure.

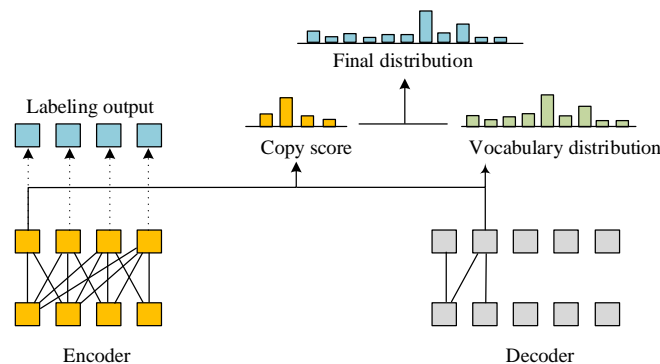


Fig. 5. Transformer structure in grammar evaluation.

Since the Transformer structure alone suffers from the problem of sparse gradients, optimization methods need to be utilized to improve this problem. Here, Adaptive moment estimation (ADAM) optimization is chosen in combination with Transformer structure. This is an adaptive learning rate optimization algorithm that is commonly used to train deep neural networks. The Adam algorithm is derived by combining the advantages of Adagrad and RMSProp algorithms to dynamically adjust the learning rate and track the exponential mean of each parameter and the exponential mean of the squared values. This adaptive learning rate can be automatically adjusted during the training process to ensure that the learning rate is neither too large nor too small, improving the training efficiency and convergence speed. Compared with the traditional gradient descent method, Adam's algorithm has faster convergence speed and higher efficiency, and is widely used in the optimization of various deep learning models. Suppose the objective function is  $f(\omega)$ , then the gradient of the objective function under Adam's algorithm for the current moment parameters  $g_t$  is shown in (10).

$$g_t = \nabla f(\omega_t) \tag{10}$$

After obtaining the gradient, it is also necessary to calculate the data of first-order momentum and second-order momentum in the process, where the solution process of first-order momentum is shown in (11).

$$m_{1t} = \phi(g_1, g_2, \dots, g_t) \tag{11}$$

Equation (11) in  $m_{1t}$  is the first-order momentum. The process of solving for second-order momentum is similar to first-order momentum, and the mathematical expression of the process is shown in (12).

$$m_{2t} = \psi(g_1, g_2, \dots, g_t) \tag{12}$$

In (12), the second-order momentum is denoted by  $m_{2t}$ . At a particular moment  $t$ , the gradient solution process of the algorithm is shown in (13).

$$\mu = l \frac{m_{1t}}{\sqrt{\mu}} \tag{13}$$

In (13),  $\mu$  represents the gradient.  $l$  represents the learning rate of the algorithm. Adam's algorithm also needs to update the parameters, and the mathematical procedure of parameter update is shown in (14).

$$\omega_t = -(\mu_{t-1} - \omega_{t-1}) \tag{14}$$

In (14),  $\omega_t$  represents the parameters at the time of  $t$ . Finally, the functions  $\phi(g_1, g_2, \dots, g_t)$  and  $\psi(g_1, g_2, \dots, g_t)$  for solving the first- and second-order momentum are defined as shown in (15).

$$\begin{cases} \phi(g_1, g_2, \dots, g_t) = lm_{1,t-1} + (1-l)g_t \\ \psi(g_1, g_2, \dots, g_t) = lm_{2,t-1} + (1-l)g_t^2 \end{cases} \tag{15}$$

This completes the construction of a grammatical analysis model for English literature based on statistical language models. The complete flowchart of the proposed algorithm can be summarized in the form shown in Fig. 6. The reordering algorithm based on minimum error training is used to adjust the output of the English corpus based on the statistical language model, while the lexical indication model and lexical N-element co-occurrence model are proposed to further optimize the output of the corpus. The proposed reordering algorithm is applied to the English corpus for English literary analysis, and it can analyze the utterances more effectively and make the output results closer to natural language. Applying this feature to English literary grammar analysis, the study combines the improved Transformer structure to propose an English grammar analysis model which can analyze and point out the grammars that may be problematic in English literature. In this part, the Transformer structure improved by Adam is used to process English literary texts and analyze them based on a corpus.

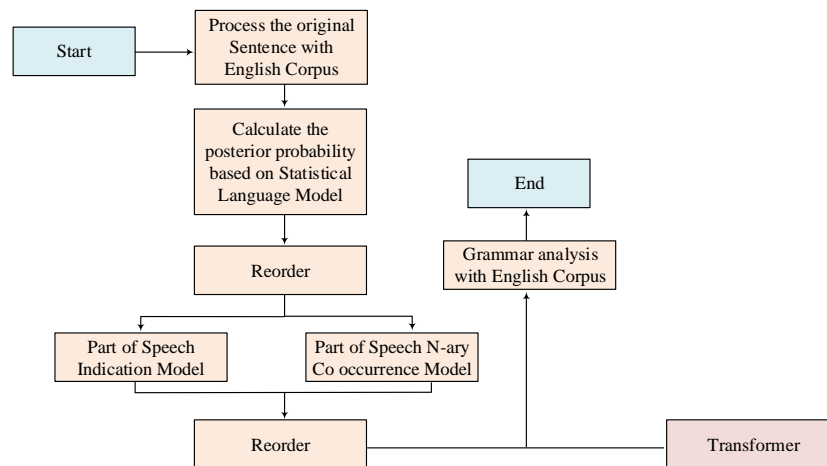


Fig. 6. Flow chart of proposed algorithm.

#### IV. MODEL TESTING AND RESULT PRESENTATION

This test focuses on the reordering algorithm for the English corpus and the grammatical analysis model of English literature combined with this algorithm. To ensure that the performance of the algorithm is fully exploited, adequate configurations as well as a large amount of data are required. The various environment configurations and the corpus used for the process of this test are shown in Table I. For system stability reasons, Windows 10 was chosen as the operating environment and Python was used as the programming environment. Four English corpora were selected, namely Gutenberg, Wikitext-103, News crawl 2018, and Tatoeba. The lowest of these databases contained 1. The lowest of these databases contains 1,000,000 statements and the highest contains 4,000,000 statements. The four databases have a total of 10,000,000 statements. The large volume of data eliminates the impact of various special cases in the experiment.

First, we measure the Perplexity of the English corpus based on the proposed reordering technique. Perplexity is an important index to evaluate the performance of linguistic statistical models, which represents the average number of branches of the target text. The reciprocal of Perplexity expresses the average probability of each word. When the language model has low Perplexity, it means that it has high performance. A high degree of Perplexity means that the model selection is more difficult and the performance is lower. Fig. 7 shows the test results of algorithm Perplexity. In order to get comparable results, N-gram algorithm and unimproved minimum error training method are used for comparison. Fig. 7(a) shows the Perplexity of several algorithms in the text with a large amount of data, and Fig. 7(b) shows their

performance in the text with a small amount of data. When the test text is a large text with a size of more than 100kb, the Perplexity of several algorithms fluctuates less. Their fluctuation range is between 400 and 550. When the text is a small file of 20kb or less, the Perplexity of several algorithms fluctuates greatly, ranging from 150 to 800. On the whole, the Perplexity of the proposed algorithm is lower than that of the other two algorithms under each TXT text, which shows that the proposed sub algorithm optimization can effectively reorder, thus controlling the complexity of the language model and ensuring the efficiency of the model.

After completing the evaluation of the perplexity, the accuracy of the algorithm output also needs to be evaluated. Since N-gram has been widely used in related fields, N-gram is directly used here as a comparison object. Fig. 8 shows the results of comparing the output accuracy of the proposed reordering algorithm with N-gram. The curves in the Fig. 8 indicate the difference in accuracy between the two on the same text, and positive values indicate that the accuracy of the proposed algorithm is higher than that of N-gram, while negative values indicate the opposite from the overall view of the curves. The majority of the accuracy curves are above 0, i.e., the proposed algorithm is more accurate than N-gram for most of the texts. The proposed algorithm is up to 14.5% more accurate than N-gram. In the few texts where its accuracy is lower than N-gram, its accuracy is no less than 5% of N-gram. A larger sample size eliminates accidental phenomena, so based on the results, although both perform negatively when dealing with different texts, the proposed algorithm has a higher reordering ability than the widely used N-gram algorithm in terms of accuracy.

TABLE I. TEST ENVIRONMENT CONFIGURATION AND CORPUS SELECTION

Item	Detail	
CPU	i5-13400f	
Memory	32 GB	
Operative System	Windows 10	
Programming Environment	Python	
Corpus	Wikitext-103	3,000,000 sentences
	Tatoeba	1,000,000 sentences
	Gutenberg	4,000,000 sentences
	News crawl 2018	2,000,000 sentences

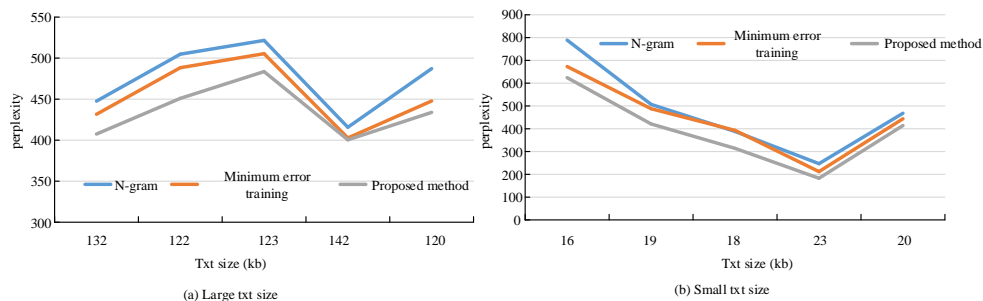


Fig. 7. The degree of confusion of the algorithm in different environments.

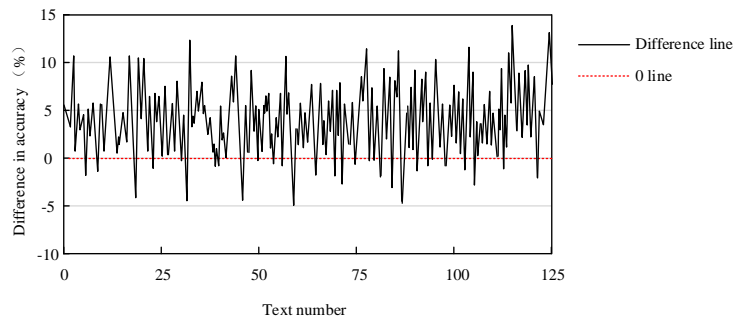


Fig. 8. Accuracy difference between the proposed algorithm and N-gram.

In addition to the accuracy, the accuracy, recall and F0.5 values of different algorithms were also compared on the dataset and the results are shown in Table II. This test was performed on the accuracy, recall and F0.5 values of each N-gram, minimum error training and the proposed reordering algorithm. The tests were done on each of the four datasets to ensure the comprehensiveness of the results. On the Gutenberg dataset, the proposed algorithm has a precision rate of 57.98 and accuracy and F0.5 values of 25.68 and 52.23, which are higher than the other two algorithms in these three dimensions. Combining the test results on the four datasets, the proposed algorithm has the highest accuracy rate of 62.13, the highest recall rate of 37.43, and the highest F0.5 value of 54.32. The proposed algorithm consistently outperforms the N-gram and the minimum error rate training methods in several dimensions of accuracy rate, recall rate, and F0.5 value, both in terms of individual dataset comparisons and in terms of the dataset as a whole.

After completing the analysis of the proposed reordering algorithm, the testing of the English literary grammar analysis algorithm based on this algorithm is continued. Since grammatical analysis mainly deals with natural language,

human analysis from experts is currently the most correct way for natural language processing. Therefore, 750 texts were selected for the test and the results of human analysis from experts were compared with the results of the algorithm, and the results are shown in Fig. 9. The horizontal coordinates in this Fig. 9 represent the different texts and the vertical coordinates represent the scores of the two methods for the grammar. Looking at the overall distribution of scores, we can see that the distribution of scores scored by the algorithm is more concentrated than that scored by the expert human, but in general there is a certain correspondence. The expert scores are concentrated in the range of 98 to 75, while the algorithmic scores are concentrated in the range of 75 to 87. The mean score of expert scoring was 85.15 and the mean score of algorithmic scoring was 84.27. The correlation analysis of the results showed that the correlation coefficient of the two scoring methods was 0.7893, which means that there is a significant correlation between them. The change of the results indicates that the proposed English grammar analysis algorithm is somewhat synchronized with the results of the human analysis, and therefore its correctness is to some extent trustworthy.

TABLE II. COMPARISON RESULTS OF ALGORITHM PERFORMANCE

Corpus	Algorithms	Precision	Recall	F0.5
Wikitext-103	Proposed	66.54	37.43	38.42
	Minimum error training	60.78	32.84	33.43
	N-gram	57.31	30.11	29.75
Tatoeba	Proposed	60.84	23.52	54.32
	Minimum error training	53.13	20.18	43.81
	N-gram	51.27	18.60	41.58
Gutenberg	Proposed	57.98	25.68	52.23
	Minimum error training	50.55	21.14	47.64
	N-gram	47.83	18.93	42.41
News crawl 2018	Proposed	62.13	27.61	46.58
	Minimum error training	57.64	24.33	40.77
	N-gram	55.53	20.58	36.12

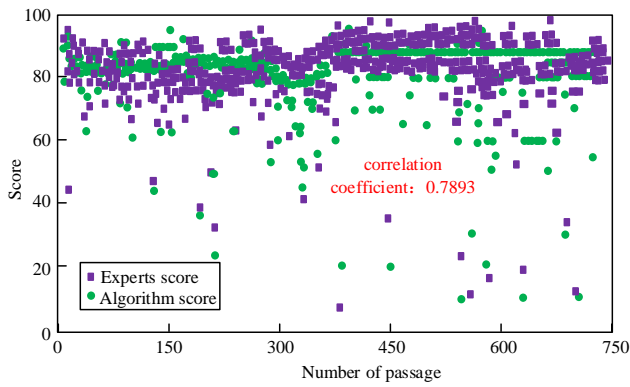


Fig. 9. Comparison results of algorithm and manual analysis.

There are currently grammar analysis algorithms being applied, and to confirm the superiority of the proposed algorithms compared to existing algorithms, a certain online

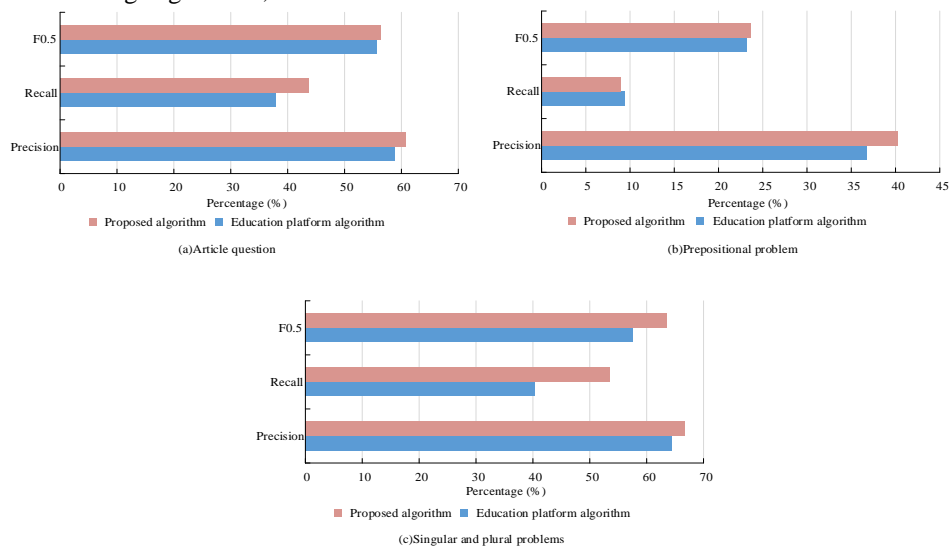


Fig. 10. Grammar problem test results.

## V. DISCUSSION

The proposed model is an English natural language analysis model based on an English corpus, designed to analyze English corpora and literature. The reordering algorithm based on Minimum Error Training is employed to adjust the output of the English corpus using statistical language models. Additionally, the introduction of the part-of-speech indicator model and part-of-speech n-gram co-occurrence model further enhances the optimization of the corpus output. When applied to English literary analysis using the proposed reordering algorithm, it facilitates more effective sentence analysis, resulting in output that closely aligns with natural language. By incorporating this feature into English literary grammar analysis, a research study proposes an improved Transformer-based English grammar analysis model to identify potential grammar issues in English literary works. In this study, an enhanced Transformer structure, utilizing improvements from Adam optimization, is utilized to process English literary texts and perform analysis based on the corpus.

teaching platform's analysis algorithm was used as the comparison object. Three common types of grammar problems were used as the comparison objects: article questions, prepositional problems, and singular and plural problems. The test results are shown in Fig. 10, where Fig. 10(a) shows the comparison results of article questions, Fig. 10(b) shows the comparison results of prepositional problems, and Fig. 10(c) shows the comparison results of singular and plural problems. Compared with the education platform algorithm, the proposed algorithm is superior in precision, recall, and F0.5 in all three dimensions. The proposed algorithm achieved a precision rate of 64.37%, a recall rate of 40.32%, and an F0.5 value of 57.51% in singular and plural problems. For article questions, the precision rate of the proposed algorithm reached 60.79%, while the education platform algorithm only reached 58.82%. Through a comprehensive analysis of the comparison results, it can be seen that the proposed grammar analysis algorithm has a stable advantage over existing algorithms.

In the results display section, multiple datasets were used to compare the proposed model with other similar models. The reason for using multiple datasets is that this comparison method can to some extent eliminate randomness and increase the reliability of experimental results. According to the experimental results, the proposed model has the highest accuracy of 62.13, the highest recall rate of 37.43, and the highest F0.5 value of 54.32 on the four datasets used. From these indicators, the proposed model has stable advantages compared to similar algorithms. Due to the fact that manual analysis by humans is currently difficult for machines to replace in the field of natural language analysis, the results of expert human analysis are also entered here and compared with the results of algorithm analysis. After conducting correlation analysis on the statistical results, it was found that the correlation coefficient between the two analysis methods was 0.7893, indicating that the results of algorithm analysis and manual analysis are to some extent similar. This means that the proposed model is to some extent close to people's processing ability of English literature and natural language.

## VI. CONCLUSION

Aiming at the optimization of current statistical language models and English corpora, as well as the gaps in automatic algorithms in the field of English literature analysis, this research proposes an improved re-sorting algorithm based on the minimum error rate training. Based on the re-sorting algorithm, a grammar analysis model for English literature is also proposed. The test results show that in the vast majority of texts, the accuracy of the proposed algorithm is higher than that of the N-gram algorithm. The proposed algorithm has a maximum accuracy of 14.5% higher than N-gram. In a small portion of text with accuracy lower than N-gram, its accuracy is not less than 5% of N-gram. On the Gutenberg dataset, the accuracy of the proposed algorithm is 57.98, with accuracy and F0.5 values of 25.68 and 52.23, which are higher than the other two comparative algorithms in these three dimensions. In addition, in terms of grammar analysis models, the correlation coefficient between model scoring and expert manpower scoring results is 0.7893, indicating a significant correlation between the two. On Singular and plural problems, the accuracy of the model's scoring reached 64.37, the recall rate was 40.32, and the F0.5 value was 57.51, all higher than existing grammar analysis models. The results show that the proposed model has considerable application potential in the field of English literature analysis.

## VII. LIMITATIONS AND FUTURE WORK

This study has made certain contributions to relevant fields, but the research results still have limitations. The proposed algorithm is greatly influenced by the size of the text. When the text is too small, there will be significant fluctuations in the performance of the model. How to maintain stable performance of algorithms at any text size is the direction of future work. In addition, this study did not focus on the consumption of algorithms, so it is necessary to evaluate this aspect in future work to determine its practical value.

## REFERENCES

- [1] Fang H, Shi H, and Zhang J, "Heuristic bilingual graph corpus network to improve English instruction methodology based on statistical translation approach," *Transactions on Asian and Low-Resource Language Information Processing*, vol. 20, no. 3, pp. 304-318, 2021.
- [2] Zhang L, and Xuan B, "Neural mechanisms and time course of the age-related word frequency effect in language production," *Advances in Psychological Science*, vol. 30, no. 2, pp. 333-342, 2022.
- [3] Niesen M, Vander Ghinst M, Bourguignon M, et al. "Tracking the effects of top-down attention on word discrimination using frequency-tagged neuromagnetic responses," *Journal of Cognitive Neuroscience*, vol. 32, no. 5, pp. 877-888, 2020.
- [4] Wei Z, and Zhang X, "A filtering algorithm of main word frequency for online commodity subject classification in e-commerce," *International Journal of Circuits*, vol. 15, no. 1, pp. 218-224, 2021.
- [5] Chaouch-Orozco A, Alonso J G, and Rothman J, "Individual differences in bilingual word recognition: the role of experiential factors and word frequency in cross-language lexical priming," *Applied Psycholinguistics*, vol. 42, no. 2, pp. 447-474, 2020.
- [6] Desai RH, Choi W, and Henderson JM. "Word frequency effects in naturalistic reading," *Language, cognition and neuroscience*, vol. 35, no. 5, pp. 583-594, 2020.
- [7] Teks P, Lampung B, Nyo D, et al. "Translation of the Lampung language text dialect of Nyo into the Indonesian language with DMT and SMT approach," *INTENSIF Jurnal Ilmiah Penelitian dan Penerapan Teknologi Sistem Informasi*, vol. 5, no. 1, pp. 58-71, 2021.
- [8] Sreelekha S, and Bhattacharyya P, "Indowordnet's help in Indian language machine translation," *AI & SOCIETY*, vol. 35, no. 1, pp. 689-698, 2020.
- [9] Collins G, Lundine J P, and Kaizar E, "Bayesian generalized linear mixed-model analysis of language samples: detecting patterns in expository and narrative discourse of adolescents with traumatic brain injury," *Journal of Speech Language and Hearing Research*, vol. 64, no. 4, pp. 1256-1270, 2021.
- [10] Ycel Z, Supitayakul P, Monden A, et al. "An Algorithm for Automatic Collation of Vocabulary Decks Based on Word Frequency," *IEICE Transactions on Information and Systems*, vol. 103, no. 8, pp. 1865-1874, 2020.
- [11] Ge Y, "The linguocultural concept based on word frequency: correlation, differentiation, and cross-cultural comparison," *Interdisciplinary Science Reviews: ISR*, vol. 47, no. 1, pp. 3-17, 2022.
- [12] Poncelas A, Wenniger G, and Way A, "Improved feature decay algorithms for statistical machine translation," *Natural Language Engineering*, vol. 28, no. 1, pp. 71-91, 2020.
- [13] Atici, Ramazan, Pala, et al. "Prediction of the ionospheric foF2 parameter using R Language forecast hybrid model library convenient time series functions," vol. 122, no. 4, pp. 3293-3312, 2022.
- [14] Guirong B, Shizhu H, Kang L, and Jun Z, "Using Pre-trained Language Model to Enhance Active Learning for Sentence Matching," *ACM transactions on Asian and Low-Resource Language Information Processing*, vol. 21, no. 2, pp. 19, 2022.
- [15] Lee O, "Statistical learning and language: English RCs and number agreement," *Studies in Linguistics*, vol. 58, pp. 251-274, 2021.
- [16] Boussakssou M, Ezzikouri H, and Erritali M, "Chatbot in Arabic language using seq to seq model," vol. 81, no. 2, pp. 2859-2871, 2022.
- [17] Ming Y, and Yi P, "Meta-learning for compressed language model: A multiple choice question answering study," *Neurocomputing*, vol. 487, pp. 181-189, 2022.
- [18] Ivan F, Alexey Z, Pavel B, Ekaterina D, Nikita K, Andrey K, Ekaterina A, Evgenia K, and Evgeny B, "A differentiable language model adversarial attack on text classifiers," vol. 10, pp. 17966-17976, 2022.
- [19] Gaeta L, and Brydges C, "An examination of effect sizes and statistical power in speech, language, and hearing research," *Journal of speech, language, and hearing research: JSLHR*, vol. 63, no. 5, pp. 1572-1580, 2020.

# Marginal Distribution Algorithm for Feature Model Test Configuration Generation

Mohd Zanes Sahid, Mohd Zainuri Saringat, Mohd Hamdi Irwan Hamzah, Nurezayana Zainal

Faculty of Computer Science and Information Technology,  
Universiti Tun Hussein Onn Malaysia (UTHM), Johor, Malaysia

**Abstract**—Generating test configuration for Software Product Line (SPL) is difficult, due to the exponential effect of feature combination. Pairwise testing can generate test input for a single software product that deviates from exhaustive testing, nevertheless proven to be effective. In the context of SPL testing, to generate minimal test configuration that maximizes pairwise coverage is not trivial, especially when dealing with a huge number of features and when constraints must be satisfied, which is the case in most SPL systems. In this paper, we propose an estimation of distribution algorithm, based on pairwise testing, to alleviate this problem. Comparisons are made against a greedy-based and a constraint handling based approach. The experiments demonstrate the feasibility of the proposed algorithm, such that it achieves better test configurations dissimilarity and at the same time maintain the test configuration size and pairwise coverage. This is supported by analysis using descriptive statistics.

**Keywords**—Estimation of distribution algorithm; marginal distribution algorithm; test configuration generation; pairwise testing; software product line

## I. INTRODUCTION

Many software products developed for various domains carries some similar functionality. This software shares similar functionalities since they have been developed based on the same kind of input and output types. The similarity in the internal program structure due to identical user requirements also contributes to the commonalities among these software products. Because of this scenario, and based on the benefit of reuse principles, Software Product Line (SPL) has been developed as a software development paradigm to produce software inspired by product line approach. Developing an SPL system enables us to create a software structure that is customizable to various needs, by maximizing software artefacts reusability [1]. Due to the highly variable and reusable nature of SPL artefacts, it is uneconomic to develop software based on distinct requirements separately, as some of the functionalities are similar. However, it is difficult to employ single product development paradigm to build various software products that fulfil the needs of diverse users of a similar domain.

A unit of system function in an SPL is represented as a feature, and explicitly defined as common or variable features and utilized throughout the SPL development process. One way to model the commonalities and variabilities in an SPL is using a Feature Model (FM), based on feature modeling technique [1]. Two or more features are combined and utilized

together in a single software product. This is known as feature configuration. The flexibility of feature configuration process could result in unspecified and unintended system behavior. This might lead to incorrect execution [1]. Hence, it is crucial to test all possible feature configurations to reduce the potential misbehavior of interacting features. But, to test all possible feature configurations is unfeasible. The number of feature configurations increases dramatically as the number of features increased, making full testing of feature configurations especially in large-scale FM impractical [2], [3]. In view of this, a number of techniques have been proposed to reduce the combinatorial explosion of feature configuration testing [4] that leveraged the potential of search-based techniques. More on this is presented in Section II.

In conventional meta-heuristics approaches, probabilities are implicitly employed in the selection and re-production operators, such as mutation operator, to produce offspring [5], [6]. We identified a research gap in feature configuration exhaustive testing, such that, one can explicitly build a probabilistic model of features distribution from an initial set of test configurations. This probabilistic model allows us to estimate the distribution of highly fit features and guide us in generating subsequent candidate solutions that maximize pairwise coverage. Towards that, in Section III we strategize the test configuration generation process, and our contributions are as follows:

- 1) We devise a set of algorithms based on bivariate marginal distribution in SPL context. This approach is perceived as a lightweight variant of estimation of distribution algorithm, in which only the statistics of the population are maintained across generation, instead of the actual population.
- 2) We introduce the notion of feature configuration dependency graph in part B of Section III, which contains the dependency information between pairs of features, extracted using statistical computation.
- 3) We implement the proposed approach using Java and conducted empirical studies. Results are reported and discussed in Sections IV and V.

## II. BACKGROUND AND RELATED WORKS

### A. Feature Model

Feature Modeling is a popular way to model SPL variability and it is by far the most reported in industry. In Feature Modeling, Feature Model (FM) notation has been developed to represent features and its dependencies [1]. The



tree representation of FM is known as Feature Diagram. It presents a feature as a node, and relationship between two features as an edge. Different types of edges can be assigned between features, which represent the relationship of type mandatory, optional, or, or alternative. Additionally, an FM might encompass some constraints which act as rules or conditions that limit the linking between features.

Fig. 1 shows a feature model named as OnlineBookstore SPL from Software Product Line Online Tools (SPLOT) [7].

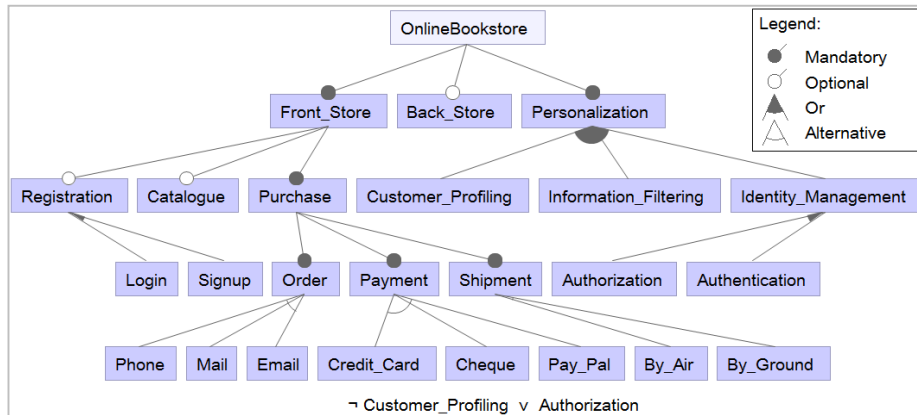


Fig. 1. A Feature Model of an OnlineBookstore SPL.

In the subsequent part of this paper, we refer each feature from our Feature Model as  $x$ . It is an integer where  $x \in [1, N]$ , having  $N$  as the maximum number of features in the FM.  $x$  can be of positive prefix (not written) to indicate the feature is included or negative prefix (explicitly written) to indicate the feature is not included.

For brevity, each feature in the feature model shown in Fig. 1 is mapped to a unique number, assigned sequentially from top to bottom, left to right. This gives us a representation as shown in Fig. 2, which is used throughout this paper.

1: OnlineBookstore	2: Front_Store
3: Back_Store	4: Personalization
5: Registration	6: Catalogue
7: Purchase	8: Customer_Profiling
9: Information_Filtering	10: Identity_Management
11: Login	12: Signup
13: Order	14: Payment
15: Shipment	16: Authorization
17: Authentication	18: Phone
19: Mail	20: Email
21: Credit_Card	22: Cheque
23: Pay_Pal	24: By_Air
25: By_Ground	

Fig. 2. Number assignment of each feature.

### B. Test Configuration

Software products of an SPL are configured and produced by combining several features. These artefacts are called feature configurations. In view of testing, test case(s) can be defined for each feature. Thus, to test a feature configuration, Test Configuration (TC), which consists of many test cases, can be generated in the same way the feature configuration is generated. A test configuration, TC, is a list of all features,

This feature model defines some of the most common features that an online bookstore system should have. It consists of 25 features, including seven mandatory features. There is one cross-tree-constraint defined, which is  $\neg$  Customer\_Profiling  $\vee$  Authorization. This constraint signifies that feature Customer\_Profiling requires the presence of feature Authorization, but not conversely. Apart from that, a couple of variation points are defined. For example, we can choose to have Signup feature apart from Login; different types of Personalization features can be incorporated; and so on.

represented by its feature number. Each feature in a test configuration can have either positive or negative prefix.

### C. Pairwise Testing

Complete testing of all possible feature configurations is not feasible. For  $n$  number of features, it requires  $2^n$  number of test configurations to cover all possible combinations, because it is either selected or excluded. This makes it exponentially proportioned to the number of features. To alleviate this obstacle, pairwise testing has been widely used as a viable solution. The ultimate goal of pairwise testing is to cover all possible pair of features at least once [8]–[10]. Thus, testing can be focused on the interaction of both features. Pairwise testing is a kind of combinatorial testing, where we choose 2 features to be considered or included in our test pool. Generalization of pairwise testing is called t-wise testing, where  $t$  indicates the number of features to choose.

For its practicality and usability, SPL pairwise testing is governed by constraints. Considering two features (1 and 2) from OnlineBookstore SPL, four pairs of tuple have to be generated, i.e. (1,2), (1,-2), (-1,2) and (-1,-2), where negative prefix indicates that the feature is not selected in the feature configuration. Due to constraints (cross-tree-constraints and relationship of features in FM), some invalid pairs will be eliminated, e.g. (-1,2) is invalid, because root feature, i.e. 1, must always be selected. The same goes with mandatory features (2, 4, 7, and so on).

If we construct one Test Configuration,  $TC_i$ , for each pair of features,  $pf_{v,w}$ , (as an example, pair of feature 1 and -5), assigned as follows;

$$pf_{1,-5} = (1, -5); \quad TC_a = [ 1, ?, ?, ?, -5, ?, ?, ?, ? ]$$

we can set any arbitrary value for other variables in  $TC_a$  (marked as ?). However, these variables could possibly be matched with other pairs of features that we should cover. Thus, if we can systematically set the values of each variable in  $TC_i$ , we could maximize the number of valid pairs in each TC so that it can minimize the number of TC.

#### D. Related Works

SPL test configuration generation techniques that are based on greedy approach or applied meta-heuristics are discussed in the first part of this section. In the second part, few selected literatures of Estimation of Distribution Algorithms (EDA), including works related to its adoption in software engineering activities are presented.

Among others, multi-objective evolutionary algorithm has been proposed for SPL testing [6]. Their motivation was to minimize the tests suite by sorting the product lines. Henard et al. [2] also employed a search-based algorithm, (1+1) Evolution Strategy (ES), to generate and prioritise covering array, guided by a (dis)similarity measure. Henard et al. mentioned that t-wise approaches for SPLs are restricted to FMs of small sizes and t-wise coverage of low strength. Both are constrained by scalability issues that result from the intractable computation for very large FMs or high t-wise strength. Therefore, they formulated the feature configuration generation problem as a search-based where the search space consists of all valid feature configurations extracted from the FM. Dissimilarity between features is used as the fitness function during the searching for better populations.

Feature configuration testing is highly influenced by the effectiveness of constraint handling techniques that eliminate invalid test configurations. One of such prominent work is published by Yu, Duan et al. [11], whereby the validity checking of test configuration is achieved using minimum invalid tuples (MITs). This approach has been implemented as a tool named LOOKUP.

Hybrid of multi-objective crow search and fruitfly optimization has been studied and offers an optimal selection of the test suites at a fairly good convergence rate [12]. Haslinger et al. [13] applied a Simulated Annealing algorithm to generate t-wise covering array and demonstrated a tool to improve the performance of SPL testing. Haslinger et al. report a speedup of over 60% on 133 publicly available feature models, while preserving the coverage of the generated tests.

Johansen et al. published their solution [3] and a tool named ICPL, which capable of processing large feature models, better execution time and produced small covering array. They used the fact that a (t-1)-wise is always a subset of the t-wise, and employed this principle to recursively build up a higher strength covering array from a smaller one.

Estimation of Distribution Algorithms (EDA) is a kind of Evolutionary Algorithms that finds near optimal solutions based on the evolution of candidate solutions satisfying some fitness functions. EDA guide the search by explicitly building the probabilistic model of promising candidate solutions. The detail discussion on EDA is beyond the scope of this paper, but interested reader can refer to papers by Ceberio et al. [14], Shirazi et al. [15], Shakya and Santana [16], Simon [17] and

Pelikan [18]. In the area of Search-Based Software Engineering (SBSE), to the best of our knowledge, no attempt has been made to employ any variant of high-order EDA (which includes bivariate or multivariate statistics) in SPL testing.

EDA have been adopted to solve many optimization problems in single software-product development such as to optimize test data generation and test suites generation [19]–[21] and refactoring [22]. The work in [19] employs bivariate EDA named as COMIT [23], in which the combination of pair of variables are viewed as tree, therefore it has a single root node. They proposed integration with data mining techniques to predict the performance of a test data generator. In the context of testing for concurrent software, detecting faults can be improved by exploiting information discovered in EDA exploration that can save future test efforts [24]. EDA has also been employed to improve software reliability prediction [25]. They reported that EDA-based approach can optimize the parameters of support vector regression in predicting the software reliability, by introducing a chaotic mutation operator into traditional EDA. Prior to that, they define the software reliability prediction problem as a combinatorial optimization problem with constraints, in which, search-based are known to be a viable solution to that problem.

### III. PROPOSED APPROACH

This section presents an evolutionary-based algorithm that generates minimal and effective SPL test configuration that satisfies pairwise coverage of features, based on bivariate marginal distribution strategy.

#### A. Marginal Distribution Algorithms of EDAs

Estimation of Distribution Algorithms (EDAs) explores the space of potential solutions following the principle of survival of the fittest of individual and populations similar to Genetic Algorithm (GA) [16], [18]. However, in EDAs, crossover and mutation operators are removed and replaced by the estimation of a probability distribution. The Probability Distribution is a model of (1) the distribution of genes across all individuals, and (2) the dependence relations or independence relations of genes between individuals.

One way to estimate the distribution of genes from all the individuals in the population is by using marginal distribution. The simplest marginal distribution calculates the probability of each candidate solutions' genes independently. This strategy is called as univariate marginal distribution. This contrasts with bivariate marginal distribution, which calculates the estimation based on the dependency of two genes. The dependency that is of our interest is the statistically significant dependency, which can be computed using Pearson's chi-square statistics [26].

The probability distribution for the univariate marginal distribution is calculated using the frequency of each gene from all or truncated individuals and stored as Probability Vector (PV). The PV will be used to sample or generate new individuals in subsequent generations. For the bivariate marginal distribution, we start with calculating the frequency of each gene. Then, we calculate the joint probability of each pair of genes, using the previously calculated frequency value. Afterward, for each pair of genes, we calculate the Pearson's chi-square tests to establish links between interdependent

genes. The result of this is a set of genes dependency and we only consider two genes as interdependent if the value is statistically significant. Next, we generate a dependency graph to store this information. The graph is acyclic and not necessarily has to be connected. All genes that have no interdependency with other genes are assigned as a root node in the graph. Whereas, for those that have a link, if both genes are not yet added to the graph, choose any gene from that pair as the root node. Add the other gene as a child node and connect them using an edge. Among them, nodes that are added earlier are called as the parent node.

Based on the constructed graph, we generate a new population. First, populate genes for root nodes using univariate frequencies. Next, for each child nodes, populate the genes using conditional probability. Perform the same process for all child nodes. From there, the standard evolutionary step is applied, which is fitness evaluation of each individual in the new population. The population is truncated, and the process repeats until an acceptable solution is found. The intuition is, univariate-based EDA is considered as a lightweight evolutionary algorithm and require small memory footprint [27], whereas the bivariate EDA manifest possible variables interdependency [18].

**B. Bivariate Distribution of SPL Features**

This section presents the mechanism and illustration of a second-order EDA (bivariate distribution) to generate pairwise test configuration for SPL. Here, we named this approach as Combinatorial Testing using Estimation of Distribution (COTED).

The proposed strategy is generally outlined as follows:

- 1) Generate a set of test configuration as the initial population. Calculate the fitness of each test configuration using the number of covered pairwise. Then, we perform truncation to select highly fit candidate solutions.
- 2) Calculate univariate frequencies and bivariate frequencies of each feature number.
- 3) Create a feature configuration dependency graph using the calculated frequency.
- 4) Generate new test configurations based on the graph.
- 5) Repeat until termination criteria are matched.

We define our fitness function as the number of pairs of features covered by each test configuration, whereby, the more pairs covered, the better the fitness. The intuition is, during the search for fitter test configurations, the stronger the dependency of a particular pair of features present in the current fittest test configuration, the more frequent it should be included in the subsequent list of test configuration. For

example, if the dependency of features 5 and 7 are statistically significant in our 10 best test configurations, we should create more test configurations using the calculated conditional distribution of features 5 and 7 in the next iteration. The definition of best test configurations refers to those that cover a higher number of pairs from our list of all valid pairs.

By modeling the non-dependency between two features in a set of test configurations, we can search for possible dependency between two features. This dependency would suggest that the two features should be paired, and those that have strong dependency should be considered first. A variant of EDA that has the capability to find this dependency is called the Bivariate Marginal Distribution Algorithm (BMEDA) [17], [28]. It uses a factorization of the univariate marginal and joint probability distribution that able to expose second-order dependencies. For our test configuration generation problem, we define the univariate marginal and joint probability distribution as follows:

**Definition 1 (univariate marginal probability)**

The probability of a feature is selected,  $p(x_i)$ , is unconditional to other features. For example,  $p(5) = 0.7$ , means that the probability of feature number 5 to be selected is 70 per cent from all test configurations.

**Definition 2 (joint probability)**

Joint probability between feature  $v$  and feature  $w$ ,  $JP_{v,w}$ , is defined as the probability of feature  $v$  and feature  $w$  been considered (either selected or not selected).

In the remaining part of this section, we elaborate the details of the mechanism to generate test configuration that satisfies pairwise coverage of feature configuration.

Step 1. Feature configurations in FM are governed by constraints, so that only valid test configurations are generated. A SAT solver is utilized to populate the seed of our search space. Once a collection of valid test configurations is available, we calculate their fitness using pairwise coverage and remove unfit test configurations. To illustrate this, we make a list of valid pair of features that needs to be covered in Listing 1. We start by populating 20 test configurations from SAT solver and calculate the number of pairwise covered by each test configuration as its fitness value. We sort and select 10 fittest test configurations as our truncated initial population, which is presented in Fig. 3.

**Pair of features, pf = { (-3,20), (-3,21), (5,19), (-5,20), (5,22), (-5,23), (-6,18), (8,9), (8,19), (8,21), (-8,23), (-9,19), (9,23), (-10,20), (-10,23), (11,19), (11,22), (-11,23), (12,19), (12,22), (-12,23), (-16,23), (19,23), (20,22), (21,24), (21,-25), (22,-24), (22,25) }**

Listing 1. Pair of valid features that needs to be covered.

Test Configuration, TC	Fitness
01: [1,2, 3,4, 5, 6,7, 8, 9, 10,-11, 12,13,14,15, 16, 17,-18, 19,-20,-21, 22,-23,-24, 25]	8
02: [1,2, 3,4, 5, 6,7, 8,-9, 10, 11,-12,13,14,15, 16,-17,-18, 19,-20,-21, 22,-23,-24, 25]	8
03: [1,2,-3,4, 5, 6,7,-8,-9, 10, 11, 12,13,14,15,-16, 17,-18,-19, 20,-21, 22,-23,-24, 25]	7
04: [1,2,-3,4, 5, 6,7, 8, 9, 10, 11,-12,13,14,15, 16, 17,-18, 19,-20,-21, 22,-23, 24,-25]	6
05: [1,2, 3,4,-5, 6,7,-8, 9, 10,-11,-12,13,14,15, 16, 17,-18,-19, 20,-21,-22, 23,-24, 25]	6
06: [1,2,-3,4, 5,-6,7,-8, 9,-10, 11, 12,13,14,15,-16,-17, 18,-19,-20,-21, 22,-23,-24, 25]	6
07: [1,2,-3,4,-5,-6,7,-8,-9, 10,-11,-12,13,14,15, 16, 17,-18,-19, 20, 21,-22,-23, 24,-25]	5
08: [1,2,-3,4,-5,-6,7,-8,-9, 10,-11,-12,13,14,15, 16,-17, 18,-19,-20,-21,-22, 23, 24,-25]	5
09: [1,2, 3,4, 5,-6,7,-8, 9, 10, 11,-12,13,14,15,-16, 17,-18,-19, 20,-21,-22, 23, 24, 25]	4
10: [1,2,-3,4, 5,-6,7,-8, 9, 10, 11,-12,13,14,15, 16,-17,-18,-19, 20,-21,-22, 23, 24,-25]	4

Fig. 3. List of initial truncated population with fitness value.

Step 2. We calculate the univariate distribution for each feature. Each feature in each test configuration has either positive or negative prefix. We compute the mean of positive number for each feature from all test configurations. These processes are presented in Algorithm 1. The result of this process is the Probability Vector, PV of our initial population, as shown in Fig. 4.

**Algorithm 1. Calculate Probability Vector**

1. Load a population of candidate solutions
2. Select 10 test configurations according to its pairwise fitness
3. Let  $n$  be the length of a test configuration
4. For  $i = 1$  to  $n$
5. Calculate the mean of positive  $i$  as  $mean\_i$
6. Set the probability vector for feature  $i$ ,  $P(i) = mean\_i$
7. Next  $i$

<b>PV:</b> [1.0,1.0,0.4,1.0,0.7,0.5,1.0,0.3,0.6,0.9,0.6,0.3, 1.0,1.0,1.0,0.7,0.6,0.2,0.3,0.5,0.1,0.5,0.4,0.4,0.6]
--

Fig. 4. Probability vector of initial population.

The next step is to calculate the joint probability of all possible value in each pair of the feature. As an example, for features 5 and 11, we calculate the occurrences of all four pairs; i.e. (5,11), (5,-11), (-5,11) and (-5,-11). Based on the population in Fig. 3, we get the joint probability value of 0.6, 0.1, 0.0 and 0.3, respectively. This process is defined in Algorithm 2, line 2 to 8.

After that, calculate the Pearson’s chi-square statistics,  $C_{v,w}$ , for each pair of features using the following equation:

$$C_{v,w} = n * \sum_{\alpha,\beta} \frac{[JP_{\alpha v, \beta w} - P(\alpha v)P(\beta w)]^2}{P(\alpha v)P(\beta w)}$$

where  $n$  is the number of test configuration

$\alpha$  is either positive or negative prefix

$\beta$  is either positive or negative prefix

For example, for  $v=5$  and  $w=11$ :

$$C_{5,11} = 10 * \left( \frac{(JP_{5,11} - P(5)P(11))^2}{P(5)P(11)} + \frac{(JP_{5,-11} - P(5)P(-11))^2}{P(5)P(-11)} + \frac{(JP_{-5,11} - P(-5)P(11))^2}{P(-5)P(11)} + \frac{(JP_{-5,-11} - P(-5)P(-11))^2}{P(-5)P(-11)} \right) = 10 * \left( \frac{(0.6 - (0.7*0.6))^2}{0.7*0.6} + \frac{(0.1 - (0.7*0.4))^2}{0.7*0.4} + \frac{(0.0 - (0.3*0.6))^2}{0.3*0.6} + \frac{(0.3 - (0.3*0.4))^2}{0.3*0.4} \right) = 6.4$$

This step is defined in Algorithm 2, line 9 to 14. Based on our sampled population, the calculated bivariate frequencies are shown in Fig. 5. Here, we are only interested in chi-square value of at least 3.84 [17], based on the degree of freedom of 1 and p value of 0.05. By calculating the chi-square values of the initial population, we choose 11 feature pairs. These pairs are conceived as having a strong dependency, due to the high degree of correlation.

**Algorithm 2. Calculate Bivariate Frequencies**

1. Initialize joint probability,  $JP$
2. For  $v = 1$  to  $n - 1$
3. For  $w = 2$  to  $n$
4. For each test configuration,  $tc$
5. Calculate joint probability between feature  $v$  and  $w$ ,  $JP_{v,w}$ , group by combination of positive and negative prefix
6. Next  $tc$
7. Next  $w$
8. Next  $v$
9. Initialize chi-square,  $C$
10. For  $v = 1$  to  $n-1$
11. For  $w = 2$  to  $n$
12. Calculate the Pearson’s chi-square statistics  $C_{v,w}$
13. Next  $w$
14. Next  $v$

Feature Pair	Chi-square
(8,19)	10.0
(24,25)	10.0
(22,23)	6.6
(5,11)	6.4
(10,18)	4.5
(3,24)	4.5
(5,22)	4.4
(8,22)	4.4
(12,22)	4.4
(8,20)	4.4
(6,8)	4.4
Other pair	<3.84

Fig. 5. Bivariate frequencies of the initial population.

Step 3. The succeeding step is to create a Feature Configuration Dependency Graph (FCDG). We define FCDG as a forest and are specified in Definition 3.

Definition 3. (Feature Configuration Dependency Graph, FCDG).

$FCDG = (V,E)$ , where  $V$  is the set of all features available in the forest, and  $E$  is the set of all edges between some ordered pairs of features. FCDG contains a collection of possibly disconnected trees.

Each feature is represented by a node, and dependency between the pair of features is represented by an edge. The dependency between features is to be calculated based on conditional probability, thus its relationship is of type directional. Therefore, we link the respective nodes in our FCDG using directed edges.

We define the following six properties for the FCDG:

- 1) The indegree of a node is the number of edges directing to that node. Each node has zero or one indegree.
- 2) The outdegree of a node is the number of edges leading away from that node. Each node has zero or more outdegree.
- 3) A node with zero indegree and non-zero outdegree is called as a root node. FCDG can have more than one root node.
- 4) A node with non-zero indegree is called as a child node.
- 5) A node with non-zero outdegree is called as a parent node.
- 6) A node without a degree is called as a standalone node.

The generation process of FCDG starts by selecting a random feature and adds it to our graph. Then, add a dependent feature by finding another feature having the highest chi-square value of at least 3.84, and add it to the graph. Repeat this step until no more features fulfil this criterion. Then, select another random feature and repeat the whole process until all features are added to the graph. This process is defined in Algorithm 3.

**Algorithm 3. Create Feature Configuration Dependency Graph**

1. Let  $W$  as the set of all features
2. Let  $F$  as an empty graph, consists of empty  $V$  and  $E$
3. Select a random feature,  $r$ , from  $W$
4. Add  $r$  to the graph,  $V$
5. Remove  $r$  from  $W$
6. If there are no more features in  $W$ , goto end
7. For each remaining features in  $W$
8. Find a feature,  $s$ , that has the highest dependency to feature  $r$
9. If found
10. add  $s$  to  $V$
11. removes  $s$  from  $W$
12. add set  $\{r,s\}$  into  $E$
13. if not found
14. goto step 3
15. end if
16. End

Executing this algorithm against the values from Fig. 5 will result in a graph with the following attributes:

- The set of all features,  $V = \{1,2,3,\dots,25\}$
- Edges between some ordered feature pairs,  $E = \{ \{3,24\}, \{5,11\}, \{8,6\}, \{8,19\}, \{8,20\}, \{18,10\}, \{22,5\}, \{22,8\}, \{22,12\}, \{22,23\}, \{24,25\} \}$

In this case, the FCDG for our running example consists of 25 nodes with 11 edges. This can be graphically presented using a forest with three disconnected trees as shown in Fig. 6. All nodes in white colour are standalone nodes, in which no dependency to other nodes is discovered. The rest (coloured nodes) are nodes with dependency. As an example, it is shown that features 18 and 10 are highly dependent. From Fig. 3, feature 18 is always negative whenever feature 10 is positive, and the other way round. Another example is between feature 8 and 19. It is of high frequency that both having negative values in the same row. The same relationship (of a certain pattern) can be observed for the rest of the pairs.

Step 4. Once we have the dependency graph, we can proceed with generating a new population. It consists of two parts, (1) to populate root nodes, and (2) to populate child node.

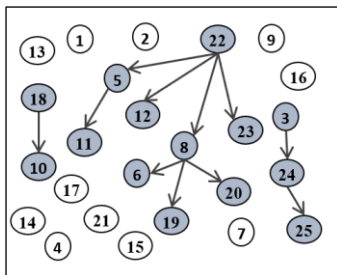


Fig. 6. The feature configuration dependency graph of the initial population.

We start by populating all features correspond to the root nodes in our graph. The features are assigned with positive or negative values using univariate probabilities. Then, we populate the remaining features that correspond to the child nodes. This is performed by calculating the conditional probabilities of the child nodes given its parent nodes. We define the conditional probabilities for our strategy as follows:

**Definition 4. (conditional probability)**

Conditional probability of feature  $s$  and feature  $r$ ,  $CP(s/r)$  is defined as the probability of feature  $s$  to be selected, given feature  $r$  been selected. It is calculated using the joint probability of  $s$  and  $r$ ,  $JP_{s,r}$ , divide by the univariate probability of  $r$ , i.e  $P(r)$ .

$$CP(s | r) = \frac{JP_{s,r}}{P(r)}$$

This process is defined from line 3 to line 8 of Algorithm 4.

**Algorithm 4. Populate New Generation**

1. For each root nodes,  $r$ , in  $G$
2. Populate new generation having positive/negative value of  $r$  using univariate frequencies
3. For other nodes,  $s$ , in  $G$
4. If parent node of  $s$  has been populated
5. Populate positive/negative value of  $s$  based on the conditional probability of  $s$  given parent of  $s$
6. If all features have been populated, goto end
7. Next root node
8. End

To demonstrate the first part, which is populating all the root nodes, Fig. 7 shows a possible assignment for 20 test configurations of our new generation. For example, for feature 16, from our initial generation, the PV value for feature 16 is 0.7, hence 70% of the new generation should have the positive value of 16. This can be achieved by using random numbers generated from a uniform distribution between 0 and 1. As per shown in Fig. 7, the outcome of this strategy is the assignment of a positive value of 16 for test configurations TC01, TC04-TC08, TC10, TC11, TC14 TC16 and TC18 TC20. The remaining test configurations are assigned with -16. We apply the same strategy to populate the remaining root node features, and values are presented in Fig. 7. For non-root node features, which we mark with unfilled squares ( $\square$ ), will be populated later.

The second part populates the remaining features, with respect to the child nodes from our dependency graph, i.e. features 5, 6, 8, 10, 11, 12, 19, 20, 23, 24, 25. Let us choose feature 12 as an example. Since feature 22 has been assigned with values, we assign feature 12 given the respective values of feature 22, using conditional distribution. It can be calculated using the joint probability of both features having positive values in the initial population, i.e. 0.3. Then divide by the probability vector of feature 22, i.e. 0.5. This equates to 0.6. Thus, we populate 60% of feature 12 with positive values for test configuration having positive 22. Similarly, calculate the probability of positive 12 given the negative value of feature 22, and use the result to populate the value of remaining test configurations. Once all values for feature 12 have been

assigned, we use the same strategy to populate the remaining features. A possible outcome of this process is shown in Fig. 8.

Step 5. The final step in this iteration is to calculate the fitness of each individual in the new generation. We count how many pairs from Listing 1 matched with each pair of features in each test configuration. The fitness values are shown in the

right column of Fig. 8. It is observed that three test configurations (TC10, TC17, and TC18) have better fitness value (marked with \*) compared to the previous generation of test configuration (refer Fig. 3). Truncation is again applied to select only ten highly fit individuals. The whole process repeats from Algorithm 1 and continue until the intended pairwise coverage has been met.

```

List of root node features=
[1, 2, 3, 4, 7, 9, 13, 14, 15, 16, 17, 18, 21, 22]
PV (for root node features)=
[1.0, 1.0, 0.4, 1.0, 1.0, 0.6, 1.0, 1.0, 1.0, 0.7, 0.6, 0.2, 0.1, 0.5]
New Generation of Test Configuration, TC:
01: [1,2,-3,4,□,□,7,□, 9,□,□,□,13,14,15, 16,-17,-18,□,□,-21,-22,□,□,□]
02: [1,2,-3,4,□,□,7,□,-9,□,□,□,13,14,15,-16, 17, 18,□,□,-21,-22,□,□,□]
03: [1,2,-3,4,□,□,7,□, 9,□,□,□,13,14,15,-16, 17,-18,□,□,-21, 22,□,□,□]
04: [1,2, 3,4,□,□,7,□,-9,□,□,□,13,14,15, 16, 17, 18,□,□,-21, 22,□,□,□]
05: [1,2, 3,4,□,□,7,□, 9,□,□,□,13,14,15, 16, 17,-18,□,□,-21,-22,□,□,□]
06: [1,2,-3,4,□,□,7,□, 9,□,□,□,13,14,15, 16, 17, 18,□,□,-21,-22,□,□,□]
07: [1,2,-3,4,□,□,7,□,-9,□,□,□,13,14,15, 16, 17,-18,□,□,-21, 22,□,□,□]
08: [1,2, 3,4,□,□,7,□, 9,□,□,□,13,14,15, 16, 17,-18,□,□,-21,-22,□,□,□]
09: [1,2,-3,4,□,□,7,□, 9,□,□,□,13,14,15,-16, 17, 18,□,□,-21,-22,□,□,□]
10: [1,2, 3,4,□,□,7,□, 9,□,□,□,13,14,15, 16, 17, 18,□,□,-21,-22,□,□,□]
11: [1,2, 3,4,□,□,7,□, 9,□,□,□,13,14,15, 16,-17, 18,□,□, 21, 22,□,□,□]
12: [1,2,-3,4,□,□,7,□, 9,□,□,□,13,14,15,-16,-17,-18,□,□,-21,-22,□,□,□]
13: [1,2,-3,4,□,□,7,□, 9,□,□,□,13,14,15,-16,-17,-18,□,□,-21, 22,□,□,□]
14: [1,2, 3,4,□,□,7,□,-9,□,□,□,13,14,15, 16, 17,-18,□,□, 21,-22,□,□,□]
15: [1,2,-3,4,□,□,7,□, 9,□,□,□,13,14,15, 16, 17,-18,□,□,-21,-22,□,□,□]
16: [1,2, 3,4,□,□,7,□, 9,□,□,□,13,14,15, 16,-17, 18,□,□,-21,-22,□,□,□]
17: [1,2,-3,4,□,□,7,□,-9,□,□,□,13,14,15,-16,-17,-18,□,□,-21, 22,□,□,□]
18: [1,2, 3,4,□,□,7,□,-9,□,□,□,13,14,15, 16, 17,-18,□,□,-21, 22,□,□,□]
19: [1,2,-3,4,□,□,7,□, 9,□,□,□,13,14,15, 16, 17,-18,□,□,-21, 22,□,□,□]
20: [1,2, 3,4,□,□,7,□,-9,□,□,□,13,14,15, 16,-17,-18,□,□,-21,-22,□,□,□]
    
```

Fig. 7. Populated root node features using univariate frequencies.

```

List of Non-Root node features =
[5, 6, 8, 10, 11, 12, 19, 20, 23, 24, 25]
New Generation of Test Configuration, TC:
01: [1,2,-3,4,-5,-6,7,-8, 9, 10,-11,-12,13,14,15, 16,-17,-18,-19,-20,-21,-22, 23, 24,-25] 5
02: [1,2,-3,4, 5,-6,7,-8,-9, 10, 11,-12,13,14,15,-16, 17, 18,-19, 20,-21,-22, 23, 24,-25] 5
03: [1,2,-3,4, 5, 6,7, 8, 9, 10, 11, 12,13,14,15,-16, 17,-18, 19,-20,-21, 22,-23, 24,-25] 8
04: [1,2, 3,4, 5, 6,7, 8,-9,-10,-11,-12,13,14,15, 16, 17, 18, 19,-20,-21, 22,-23,-24,-25] 5
05: [1,2, 3,4, 5,-6,7,-8, 9, 10, 11,-12,13,14,15, 16, 17,-18,-19,-20,-21,-22, 23,-24, 25] 3
06: [1,2,-3,4, 5, 6,7,-8, 9, 10, 11,-12,13,14,15, 16, 17, 18,-19,-20,-21,-22, 23, 24,-25] 3
07: [1,2,-3,4, 5, 6,7, 8,-9, 10, 11,-12,13,14,15, 16, 17,-18, 19,-20,-21, 22,-23, 24,-25] 6
08: [1,2, 3,4, 5,-6,7,-8, 9, 10, 11,-12,13,14,15, 16, 17,-18,-19, 20,-21,-22, 23,-24,-25] 3
09: [1,2,-3,4, 5,-6,7,-8, 9,-10,-11,-12,13,14,15,-16, 17, 18,-19,-20,-21,-22, 23, 24,-25] 7
10: [1,2, 3,4,-5,-6,7,-8, 9,-10,-11,-12,13,14,15, 16, 17, 18,-19, 20,-21,-22, 23,-24,-25] 9 *
11: [1,2, 3,4, 5, 6,7,-8, 9, 10, 11, 12,13,14,15, 16,-17, 18,-19,-20, 21, 22,-23,-24, 25] 5
12: [1,2,-3,4,-5,-6,7,-8, 9, 10, 11,-12,13,14,15,-16,-17,-18,-19, 20,-21,-22, 23, 24,-25] 7
13: [1,2,-3,4, 5, 6,7, 8, 9, 10, 11,-12,13,14,15,-16,-17,-18, 19,-20,-21, 22,-23,-24, 25] 8
14: [1,2, 3,4,-5, 6,7,-8,-9, 10,-11,-12,13,14,15, 16, 17,-18,-19, 20, 21,-22, 23,-24,-25] 6
15: [1,2,-3,4,-5, 6,7,-8, 9, 10,-11,-12,13,14,15, 16, 17,-18,-19, 20,-21,-22, 23, 24,-25] 7
16: [1,2, 3,4,-5,-6,7,-8, 9,-10, 11,-12,13,14,15, 16,-17, 18,-19, 20,-21,-22, 23,-24,-25] 8
17: [1,2,-3,4, 5, 6,7, 8,-9, 10, 11, 12,13,14,15,-16,-17,-18, 19,-20,-21, 22,-23,-24,-25] 9 *
18: [1,2, 3,4, 5, 6,7, 8,-9, 10, 11, 12,13,14,15, 16, 17,-18, 19,-20,-21, 22,-23,-24, 25] 10 *
19: [1,2,-3,4, 5, 6,7, 8, 9, 10, 11, 12,13,14,15, 16, 17,-18, 19,-20,-21, 22,-23, 24,-25] 8
20: [1,2, 3,4,-5,-6,7,-8,-9, 10,-11,-12,13,14,15, 16,-17,-18,-19,-20,-21,-22, 23,-24, 25] 4
    
```

Fig. 8. Populated non-root node features using conditional distribution and calculated fitness value.

#### IV. EXPERIMENT AND RESULTS

COTED has been implemented and executed on a set of feature models from Software Product Line Online Tools (SPLOT) [7]. The objective is to measure the efficiency and effectiveness of bivariate distribution approach based on EDA in generating test configuration satisfying pairwise testing. The comparison has been made against (1) a greedy-based approach, ICPL [3] and (2) a constraint handling approach based on the minimum-invalid-tuple strategy, LOOKUP [29].

The first part assesses the efficiency by measuring the minimum number of test configurations that the three approaches able to generate. The second part measures the

quality of the generated test configuration, in terms of the frequency of pairwise tuple, and test configuration similarity. During the experiments, 8 datasets of various sizes of constrained Feature Models (FMs) have been selected from SPLOT. COTED has been executed with the population of size 800 with truncation size 100, stagnancy count of 3 executions, maximum generations were 5000 and execution timeout of 1800 seconds.

##### A. Minimum Number of Test Configurations

This is the most used metric that evaluates the efficiency of the solution for SPL test configuration generation [12]. It calculates the number of test configurations generated using a particular approach that either fully satisfies the pairwise

coverage, or partially fulfil the coverage with a decent percentage. However, the latter does not conform to the definition of pairwise testing, i.e. to have all pairs covered at least once. Therefore, a complete pairwise coverage is often of the goal in any SPL test configuration exercise.

Fig. 9 shows that LOOKUP is the most outstanding tool in generating the most minimal test configurations. For J2EEWebArch and CocheEcologico, it outperforms the other techniques. For others, it produces an equal number of test configuration as generated by COTED, except for SPLSimulES dataset.

**B. t-wise Frequency**

This measure has been devised by Perrouin et al. [30] as the ratio between the occurrences of t-wise and the number of test configurations generated. This can be used to check whether the solution satisfies the t-wise principle, i.e., in the solution, every valid combination of t factors must be present at least once. An optimum solution consists of combination of t factors once. This, however, is hard to achieve.

Fig. 10 shows the box plots of all evaluated techniques calculated based on the median of t-wise frequencies of the generated test configurations for each benchmark datasets. In general, the average and the dispersion of the t-wise frequencies are stable for the three techniques. Most of the results show that the frequencies are maintained low, as depicted by the concentration on the low end of the scale, except for Ecommerce (Fig. 10(a)) and Billing (Fig. 10(g)) datasets. Low frequency of t-wise in the generated test configurations indicates that there are less pairwise

occurrences; hence lower the redundancy of feature of pairs. This is useful in the event of limited time and resources available for testing, which is often the case in SPL testing.

On average, as shown in Table I, the median and standard deviation ( $\sigma$ ) of the proposed techniques resides on the decent level, which is on par with the other approaches. Even though, on average, ICPL can demonstrate lower t-wise frequency (0.288), the deviation of the overall solution is worse than the rest. On the other side of the coin, LOOKUP and COTED managed to cover pairwise steadily, with low variations, on average, however, it covers higher frequency than ICPL. The differences between COTED and LOOKUP are relatively low. 50 per cent of the overall medians are equal for both techniques.

Datasets	Num. of Features	Num. of Constraints	TC Generation Techniques		
			COTED	ICPL	LOOKUP
Ecommerce	10	10	6	7	6
Cellphone	11	14	7	8	7
GraphProductLine	20	30	15	17	15
SPLSimulES	32	25	10	10	11
ArcadeGame	61	87	16	18	16
J2EEWebArch	77	86	19	18	17
Billing	88	89	13	14	13
CocheEcologico	94	131	92	93	90

Fig. 9. Minimum number of test configuration generated.

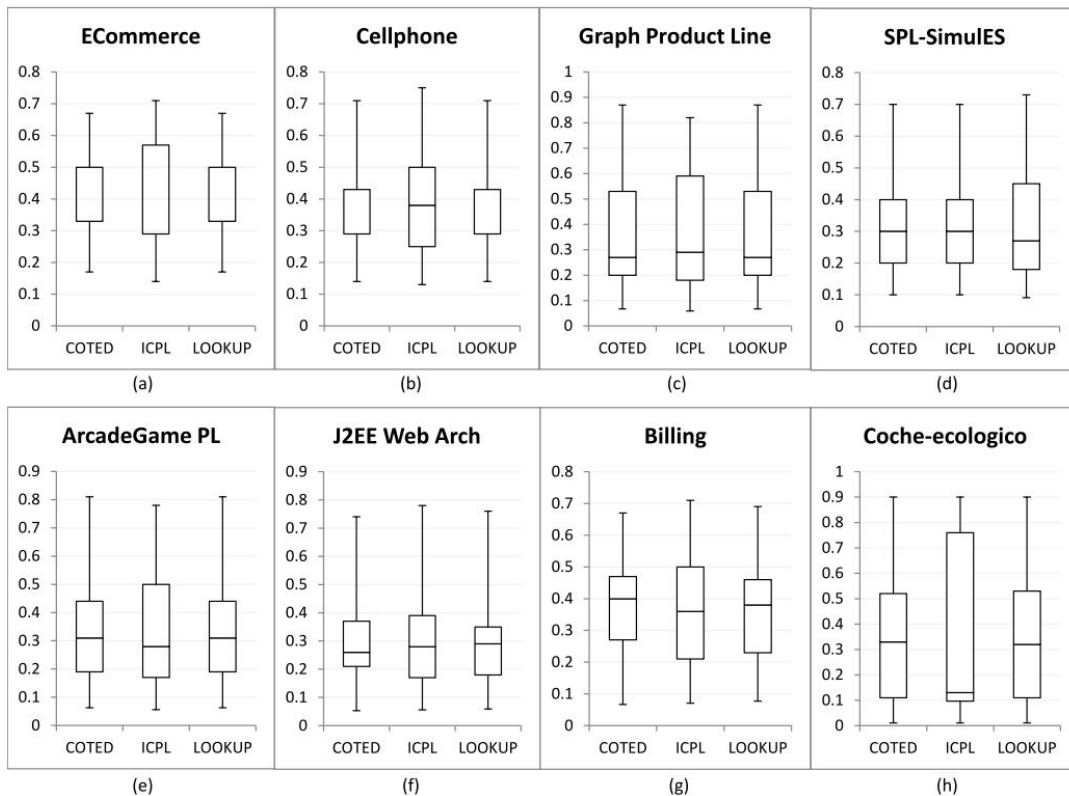


Fig. 10. Box plots for the median of t-wise frequency of the three approaches.

TABLE I. MEDIAN AND STANDARD DEVIATION ( $\sigma$ ) OF T-WISE FREQUENCY

Techniques Datasets	COTED		ICPL		LOOKUP	
	Median	$\sigma$	Median	$\sigma$	Median	$\sigma$
ECommerce	0.33	0.1634	0.29	0.1910	0.33	0.1634
Cellphone	0.29	0.1615	0.38	0.1636	0.29	0.1615
SPL-SimulES	0.3	0.1498	0.3	0.1543	0.27	0.1470
ArcadeGamePL	0.31	0.1695	0.28	0.2101	0.31	0.1771
Graph Product Line	0.27	0.2280	0.29	0.2289	0.27	0.2263
J2EE Web Arch	0.26	0.1440	0.28	0.1566	0.29	0.1391
Billing	0.4	0.1530	0.36	0.1782	0.38	0.1553
Coche-ecologico	0.33	0.2488	0.13	0.3353	0.32	0.2530
Average	0.311	0.177	0.288	0.202	0.307	0.177

C. Test Configurations Similarity

The third measurement is test configuration similarity [30]. The objective is to assess the degree of similarity between test configurations among a different set of solutions. The similarity between two test configurations is calculated using Jaccard Index, *Jac*. Given a and b as the two test configurations, we calculate *Jac*(a,b) as follows:

$$Jac(a,b) = \frac{|a \cap b|}{|a \cup b|}$$

The presence of all mandatory features is a must in all test configurations. Since all solutions from the three techniques are of valid test configurations, we omit the similarity checking for mandatory features. Only optional features are observed.

This similarity measure can be used to measure the degree of diversity of the generated solutions. Lower Jaccard Index value indicates that the test configurations are less likely to be similar, hence more diversified. Fig. 11 shows the box plots calculated based on the median of the test configuration

similarity from the generated solutions for each benchmark datasets. Overall, the averages of the test configuration similarity are low and encouraging among all techniques, and the dispersions of the median are stable for all techniques. This is depicted in Fig. 11 based on the trend of right skewness, as most medians are closer to the first quartile than the third quartile. COTED performance is on par with LOOKUP, and in fact, it managed to outperform LOOKUP at SPL-SimulES dataset. Overall, COTED and LOOKUP outperform ICPL for most datasets.

With respect to the average and measure of dispersion, as shown in Table II, LOOKUP performed better than the rest, with the exception to three datasets (Cellphone, SPL SimulES and ArcadeGamePL) where COTED has a bit lower median values. Meanwhile, the median of COTED is better than ICPL, with lower median and  $\sigma$  on five datasets (ECommerce, Cellphone, ArcadeGamePL, Graph Product Line and Coche ecologico). This suggests, on average, it produces more dissimilar sets of test configurations.

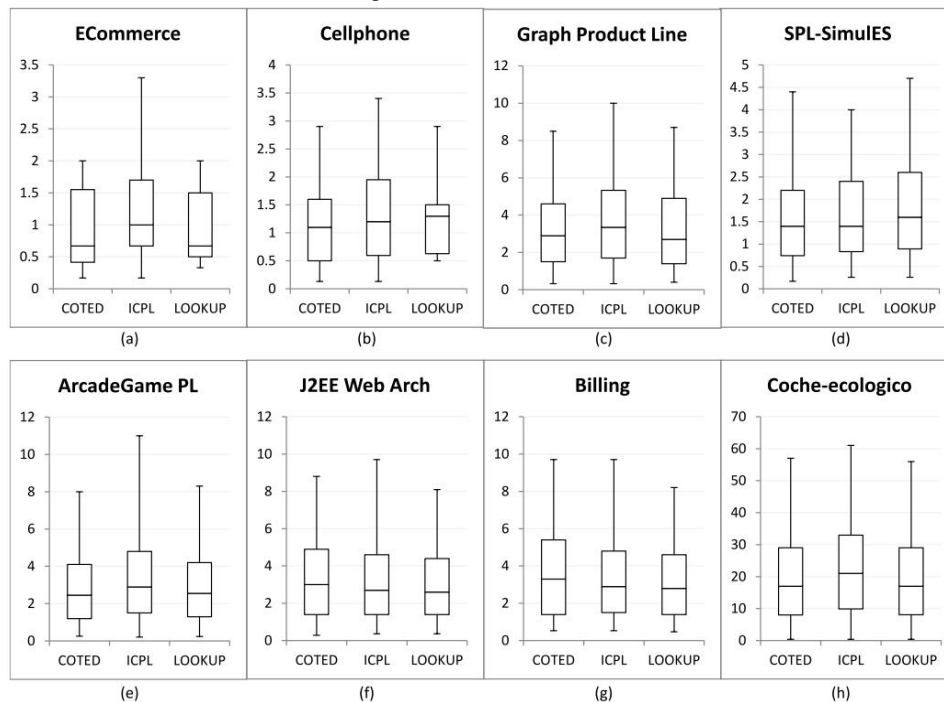


Fig. 11. Box plots for the median of t-wise similarity of the three approaches.



TABLE II. MEDIAN AND STANDARD DEVIATION ( $\sigma$ ) OF TEST CONFIGURATION SIMILARITY

Techniques Datasets	COTED		ICPL		LOOKUP	
	Median	$\sigma$	Median	$\sigma$	Median	$\sigma$
ECommerce	0.67	0.6581	1	0.8064	0.67	0.6489
Cellphone	1.1	0.8134	1.2	0.8955	1.3	0.7188
SPL-SimulES	1.4	1.0682	1.4	1.0199	1.6	1.1632
ArcadeGamePL	2.45	1.9168	2.9	2.2197	2.55	1.9063
Graph Product Line	2.9	2.1399	3.35	2.4305	2.7	2.1715
J2EE Web Arch	3	2.2429	2.7	2.1194	2.6	1.9755
Billing	3.3	2.3822	2.9	2.2185	2.8	2.0049
Coche-ecologico	17	13.6009	21	14.3204	17	13.3103
Average	3.977	3.102	4.556	3.253	3.902	2.987

## V. DISCUSSIONS

By calculating the marginal distribution between every two features in a particular sample, we can infer its connection. And based on that strong assumption, the population evolve towards more frequently connected features. This can directly be translated to more pairwise coverage. The ability to maximize pairwise coverage at each evolution cycle results in the reduction in the overall cycles of exploration, and subsequently, reduce the number of generated test configuration that fulfil pairwise coverage.

This strategy has been evaluated against two current approaches, i.e. greedy-based and minimum invalid tuple based. Of the three strategies, the minimum invalid tuple-based strategy performed the best, but, competitively challenged by COTED, and this is supported by results analysis using descriptive statistics.

Even though the performance of COTED is shown to be comparable, if not better than other approach, it provides us with a set of knowledge on the problem structure. By analysing the evolution of the probability models during test configuration generation, we discover a set of data on how the problem is being solved. We also gain knowledge on how features are distributed in the population with respect to other features. We explicitly acquire this in the form of feature configuration dependency graph which stores a set of feature pairs that have strong dependency. This information is deemed crucial as it could help us (1) decide how to prioritize the test configurations in pairwise testing, and (2) infer a higher order marginal distribution based on the collection of dependency knowledge.

As compared to test generation, previous literature highlighted that test prioritization for SPL is insufficiently researched, especially on one that is based on feature reusability [31]. Reusable features are features that appear more frequently in final software products than the others. Hence, calculating the frequency might help in extracting the most reusable one. In view of interaction testing, two interacting features are of one main concern. Thus, to find those reusable interactions could mean to find common feature interactions.

The dependency knowledge in the form of a collection of feature configuration dependency graphs are acquired iteratively from second-order probabilistic model. As opposed to computing a higher-order probabilistic model (which involves multivariate computation), this process is more viable as it incurs much lower cost. Additionally, a higher-order probabilistic model is achievable by grouping or clustering lower-order dependencies which contains highly interacting sets of variables [32]. Therefore, we could leverage a lightweight second-order iterative computation for practical higher-order computation. This remains to be investigated and thus motivate our future work.

## VI. CONCLUSION AND FUTURE WORKS

Generating efficient and effective test configurations for SPL is difficult. One way to feasibly tackle the combinatorial explosions of feature configuration testing is by leveraging pairwise testing.

Based on the work we conducted throughout this study, we found that the marginal distribution algorithm-based approach is a feasible and competitive strategy. It allows us to reduce the number of required test configuration from an exhaustive approach by leveraging pairwise coverage as its fitness function. Our proposed strategy managed to generate the solution of similar quality in terms of t-wise frequency and test configuration diversity, compared to those generated by state-of-the-art approaches. The outcome of the proposed strategy is two-fold. First, it generates minimized test configuration for pairwise testing. Secondly, the inherent ability of the strategy to extract the dependency knowledge in the form of feature configuration dependency graphs. As per our knowledge, this is the first time a combinatorial interaction testing in software product line problem is being modelled and tackled by using probability based evolutionary algorithm.

## ACKNOWLEDGMENT

This research was supported by Universiti Tun Hussein Onn Malaysia (UTHM) through Tier 1 (Vote Q103).

## REFERENCES

- [1] D. Hinterreiter, K. Feichtinger, L. Linsbauer, H. Prähofer, and P. Grünbacher, "Supporting feature model evolution by lifting code-level dependencies: A research preview," in Requirements Engineering:

- Foundation for Software Quality: 25th International Working Conference, REFSQ 2019, Essen, Germany, March 18–21, 2019, Proceedings 25, 2019, pp. 169–175.
- [2] C. Henard, M. Papadakis, G. Perrouin, J. Klein, P. Heymans, and Y. le Traon, “Bypassing the Combinatorial Explosion: Using Similarity to Generate and Prioritize T-Wise Test Configurations for Software Product Lines,” *Softw. Eng. IEEE Trans.*, vol. 40, no. 7, pp. 650–670, 2014, doi: <http://doi.org/10.1109/TSE.2014.2327020>.
- [3] M. F. Johansen, Ø. Haugen, and F. Fleurey, “An algorithm for generating t-wise covering arrays from large feature models,” in *Proceedings of the 16th International Software Product Line Conference-Volume 1*, 2012, pp. 46–55. doi: [10.1145/2362536.2362547](https://doi.org/10.1145/2362536.2362547).
- [4] A. Bajaj and O. P. Sangwan, “A systematic literature review of test case prioritization using genetic algorithms,” *IEEE Access*, vol. 7, pp. 126355–126375, 2019.
- [5] C. Henard, M. Papadakis, M. Harman, and Y. Le Traon, “Combining multi-objective search and constraint solving for configuring large software product lines,” in *2015 IEEE/ACM 37th IEEE International Conference on Software Engineering*, 2015, vol. 1, pp. 517–528.
- [6] N. Khoshniat, A. Jamarani, A. Ahmadzadeh, M. Haghi Kashani, and E. Mahdipour, “Nature-inspired metaheuristic methods in software testing,” *Soft Comput.*, pp. 1–42, 2023.
- [7] M. Mendonca, M. Branco, and D. Cowan, “SPLIT: software product lines online tools,” in *Proceedings of the 24th ACM SIGPLAN conference companion on Object oriented programming systems languages and applications*, 2009, pp. 761–762.
- [8] D. M. Cohen, S. R. Dalal, J. Parelius, and G. C. Patton, “The combinatorial design approach to automatic test generation,” *IEEE Softw.*, vol. 13, no. 5, p. 83, 1996.
- [9] D. Gupta and L. Sharma, “Improved Combinatorial Algorithms Test for Pairwise Testing Used for Testing Data Generation in Big Data Applications,” in *Artificial Intelligence*, Chapman and Hall/CRC, 2021, pp. 81–90.
- [10] J. Ferrer, F. Chicano, and J. A. Ortega-Toro, “CMSA algorithm for solving the prioritized pairwise test data generation problem in software product lines,” *J. Heuristics*, vol. 27, pp. 229–249, 2021.
- [11] L. Yu, F. Duan, Y. Lei, R. N. Kacker, and D. R. Kuhn, “Combinatorial Test Generation for Software Product Lines Using Minimum Invalid Tuples,” in *15th International Symposium on High-Assurance Systems Engineering (HASE)*, 2014, pp. 65–72. doi: [10.1109/HASE.2014.18](https://doi.org/10.1109/HASE.2014.18).
- [12] P. Ramgouda and V. Chandraprakash, “Constraints handling in combinatorial interaction testing using multi-objective crow search and fruitfly optimization,” *Soft Comput.*, vol. 23, no. 8, pp. 2713–2726, 2019, doi: [10.1007/s00500-019-03795-w](https://doi.org/10.1007/s00500-019-03795-w).
- [13] E. N. Haslinger, R. E. Lopez-Herrejon, and A. Egyed, “Improving CASA runtime performance by exploiting basic feature model analysis,” *arXiv Prepr. arXiv1311.7313*, 2013.
- [14] J. Ceberio, A. Mendiburu, and J. A. Lozano, “A roadmap for solving optimization problems with estimation of distribution algorithms,” *Nat. Comput.*, pp. 1–15, 2022.
- [15] A. Shirazi, J. Ceberio, and J. A. Lozano, “EDA++: Estimation of distribution algorithms with feasibility conserving mechanisms for constrained continuous optimization,” *IEEE Trans. Evol. Comput.*, vol. 26, no. 5, pp. 1144–1156, 2022.
- [16] S. Shakya and R. Santana, “A Review of Estimation of Distribution Algorithms and Markov Networks,” in *Markov Networks in Evolutionary Computation*, vol. 14, S. Shakya and R. Santana, Eds. Springer Berlin Heidelberg, 2012, pp. 21–37. doi: [10.1007/978-3-642-28900-2\\_2](https://doi.org/10.1007/978-3-642-28900-2_2).
- [17] D. Simon, “Estimation of Distribution Algorithms,” in *Evolutionary Optimization Algorithms*, John Wiley & Sons, 2013, pp. 313–347.
- [18] M. Pelikan, M. Hauschild, and F. Lobo, “Estimation of Distribution Algorithms,” in *Springer Handbook of Computational Intelligence*, J. Kacprzyk and W. Pedrycz, Eds. Springer Berlin Heidelberg, 2015, pp. 899–928. doi: [10.1007/978-3-662-43505-2\\_45](https://doi.org/10.1007/978-3-662-43505-2_45).
- [19] R. Sagarna and J. Lozano, “Software Metrics Mining to Predict the Performance of Estimation of Distribution Algorithms in Test Data Generation,” in *Knowledge-Driven Computing*, vol. 102, C. Cotta, S. Reich, R. Schaefer, and A. Ligeza, Eds. Springer Berlin Heidelberg, 2008, pp. 235–254. doi: [10.1007/978-3-540-77475-4\\_15](https://doi.org/10.1007/978-3-540-77475-4_15).
- [20] R. Sagarna, A. Arcuri, and Y. Xin, “Estimation of distribution algorithms for testing object oriented software,” in *Evolutionary Computation, 2007. CEC 2007. IEEE Congress on*, 2007, pp. 438–444. doi: [10.1109/cec.2007.4424504](https://doi.org/10.1109/cec.2007.4424504).
- [21] R. Sagarna and J. A. Lozano, “Scatter Search in software testing, comparison and collaboration with Estimation of Distribution Algorithms,” *Eur. J. Oper. Res.*, vol. 169, no. 2, pp. 392–412, 2006, doi: <http://dx.doi.org/10.1016/j.ejor.2004.08.006>.
- [22] N. Sadat Jalali, H. Izadkhah, and S. Lotfi, “Multi-objective search-based software modularization: structural and non-structural features,” *Soft Comput.*, vol. 23, no. 21, pp. 11141–11165, 2019, doi: [10.1007/s00500-018-3666-z](https://doi.org/10.1007/s00500-018-3666-z).
- [23] S. Baluja and S. Davies, “Fast probabilistic modeling for combinatorial optimization,” in *AAAI/IAAI*, 1998, pp. 469–476.
- [24] J. Staunton and J. Clark, “Applications of Model Reuse When Using Estimation of Distribution Algorithms to Test Concurrent Software,” in *Search Based Software Engineering*, vol. 6956, M. Cohen and M. Ó Cinnéide, Eds. Springer Berlin Heidelberg, 2011, pp. 97–111. doi: [10.1007/978-3-642-23716-4\\_12](https://doi.org/10.1007/978-3-642-23716-4_12).
- [25] C. Jin and S.-W. Jin, “Software reliability prediction model based on support vector regression with improved estimation of distribution algorithms,” *Appl. Soft Comput.*, vol. 15, pp. 113–120, 2014, doi: <http://dx.doi.org/10.1016/j.asoc.2013.10.016>.
- [26] M. Pelikan and H. Mühlenbein, “Marginal distributions in evolutionary algorithms,” in *Proceedings of the International Conference on Genetic Algorithms Mendel*, 1998, pp. 90–95.
- [27] M. Hauschild and M. Pelikan, “An introduction and survey of estimation of distribution algorithms,” *Swarm Evol. Comput.*, vol. 1, no. 3, pp. 111–128, 2011, doi: <http://dx.doi.org/10.1016/j.swevo.2011.08.003>.
- [28] M. Pelikan and H. Mühlenbein, “The bivariate marginal distribution algorithm,” in *Advances in Soft Computing*, Springer, 1999, pp. 521–535.
- [29] L. Yu, F. Duan, Y. Lei, R. N. Kacker, and D. R. Kuhn, “Combinatorial test generation for software product lines using minimum invalid tuples,” in *High-Assurance Systems Engineering (HASE)*, 2014 IEEE 15th International Symposium on, 2014, pp. 65–72.
- [30] G. Perrouin, S. Oster, S. Sen, J. Klein, B. Baudry, and Y. Le Traon, “Pairwise testing for software product lines: comparison of two approaches,” *Softw. Qual. J.*, vol. 20, no. 3–4, pp. 605–643, 2012, doi: [10.1007/s11219-011-9160-9](https://doi.org/10.1007/s11219-011-9160-9).
- [31] M. Z. Sahid, A. B. M. Sultan, A. A. Ghani, and S. Baharom, “Combinatorial Interaction Testing of Software Product Lines: A Mapping Study,” *J. Comput. Sci.*, vol. 12, no. 8, pp. 379–398, 2016, doi: <http://dx.doi.org/10.3844/jcssp.2016.379.398>.
- [32] R. Santana, P. Larranaga, and J. A. Lozano, “Learning factorizations in estimation of distribution algorithms using affinity propagation,” *Evol. Comput.*, vol. 18, no. 4, pp. 515–546, 2010.

# A QoS-Aware Resource Allocation Method for Internet of Things using Ant Colony Optimization Algorithm and Tabu Search

Shuling YIN<sup>1</sup>, Renping YU<sup>2</sup>

Hubei Open University, Wuhan 430074, China<sup>1</sup>  
Hubei Huazhong Electric Power Technology Development Co., Ltd.<sup>2</sup>

**Abstract**—In today's computing era, the Internet of Things (IoT) stands out for its implementation of automation, high-quality ecosystems, creative and efficient services, and higher productivity. IoT has found applications in various fields, such as education, healthcare, agriculture, military, and industry, where diverse resource requirements present a major challenge. To address this issue, we propose a novel QoS-aware resource allocation method for IoT systems. Our approach combines the Ant Colony Optimization (ACO) and Tabu Search (TS) algorithms to manage resources effectively, minimize energy consumption, reduce communication delays, and enhance overall system performance. Experimental results demonstrate the efficiency and effectiveness of our approach, with significant improvements in QoS metrics compared to traditional methods. By merging ACO and TS algorithms, our research contributes to the advancement of IoT capabilities, energy conservation, and business optimization.

**Keywords**—Internet of things; resource allocation; virtualization; Ant Colony Optimization; Tabu Search

## I. INTRODUCTION

The Internet of Things (IoT) enables the integration of the virtual and physical worlds, facilitating communication between various devices without human intervention [1]. This has led to a growing interest in IoT research due to its ability to enable intelligent and ubiquitous services through data aggregation, processing, analysis, and mining [2, 3]. However, the performance of IoT systems is influenced by several factors and resources, such as user requirements, energy consumption, diverse applications, storage capacity, communication needs, network bandwidth, and computing power. These resources are heterogeneous in nature, meaning they vary in their capabilities and characteristics [4]. IoT networks face resource allocation challenges, particularly in networks with heterogeneous properties. Resource allocation involves effectively managing and allocating limited resources to achieve optimal objectives [5]. The resources in IoT networks are divided into two categories: node resources and channel resources. Node resources, also known as physical resources, include storage, computational power, and energy resources. On the other hand, channel resources pertain to communication channels and networks, encompassing aspects such as channel bandwidth, load balancing, and traffic analysis [6].

Resource allocation refers to the task of efficiently assigning available resources to complete a set of tasks while

considering specific conditions and constraints [7]. The target is to optimize resource utilization and enhance the performance of the IoT platform. The resources in IoT devices are often limited due to factors such as energy constraints, processing power, and storage capacity [8]. However, IoT devices have the potential to provide various services and functionalities. Efficient resource allocation is crucial for optimizing the utilization of these limited resources and ensuring that tasks are completed effectively [9]. The heterogeneous and distributed characteristics of the devices and resources complicate IoT resource allocation. IoT devices come in different types with varying capabilities and characteristics. They may have different energy levels, processing capacities, and storage capacities. The resource allocation algorithm needs to consider these differences and allocate resources accordingly [10].

Integrating machine learning, deep learning, Artificial Intelligence (AI), and urban public transportation systems plays a pivotal role in efficiently allocating resources in the IoT. These technologies collectively form the backbone of intelligent resource management in urban environments. Machine learning algorithms enable IoT systems to adapt and optimize resource allocation strategies by analyzing vast amounts of data [11, 12]. Deep learning, a subset of machine learning, excels in pattern recognition and feature extraction, making it invaluable for understanding complex urban dynamics [13, 14]. AI systems bring decision-making capabilities to IoT devices, allowing them to dynamically adjust resource allocation based on real-time conditions and user preferences [15, 16]. Urban public transportation systems are a prime example of IoT in action, encompassing connected vehicles, smart traffic management, and passenger information systems. By leveraging the data generated within these transportation networks, machine learning models can predict traffic patterns, optimize routes, and enhance energy efficiency, leading to reduced congestion and environmental impact [17].

In recent years, researchers have turned to nature-inspired meta-heuristic algorithms to tackle optimization problems in various domains, including IoT resource allocation. These algorithms mimic natural phenomena and use techniques such as solution perturbations and stochasticity to avoid local optima and achieve optimal or near-optimal solutions. Meta-heuristic optimization algorithms have gained popularity due to their ability to handle various applications and optimization problems. These meta-heuristic algorithms can be applied to IoT resource allocation problems to meet various objectives,

such as energy efficiency, bandwidth utilization, and task allocation. Previous research efforts have explored different optimization techniques for resource allocation in IoT systems. Genetic, ACO, and Particle Swarm Optimization (PSO) algorithms are some of the methods applied to address this challenge. However, these approaches often suffer from slow convergence or suboptimal solutions, particularly when dealing with complex, combinatorial optimization problems that arise in IoT resource allocation scenarios. Therefore, it is necessary to develop more efficient and effective algorithms to cope with the complexity and variability of IoT environments.

This paper proposes a novel QoS-aware resource allocation method for IoT systems that utilizes a hybrid approach that incorporates both the Ant Colony Optimization (ACO) and the Tabu Search (TS) algorithms. The ACO algorithm is derived from the foraging behavior of ants and effectively solves combinatorial optimization problems. It employs a population of artificial ants to construct solutions by probabilistically selecting resources based on pheromone trails and heuristics. On the other hand, the TS algorithm is a local search-based metaheuristic that intensifies the search process by maintaining a tabu list, preventing revisiting previously visited solutions and encouraging the exploration of new solutions. By combining these two powerful optimization approaches, we aim to overcome the limitations of conventional resource allocation methods and provide a more efficient and effective solution regarding QoS-aware resource allocation in IoT systems.

In this context, our motivation for this research lies in the critical need for efficient resource allocation in IoT systems. Managing limited resources becomes paramount as IoT applications continue to increase across diverse domains such as education, healthcare, agriculture, military, and industry. IoT systems rely on a multitude of resources, including but not limited to storage, computational power, and energy. These resources are heterogeneous and often constrained, posing significant challenges to resource allocation. Our proposed approach addresses these challenges by prioritizing Quality of Service (QoS) in resource allocation. The potential benefits of our research are multifaceted. By effectively managing resources, our method aims to minimize energy consumption, reduce communication delays, and enhance overall system performance. This not only leads to improved operational efficiency but also contributes to sustainability efforts by reducing energy usage. Furthermore, our approach holds the promise of enabling more reliable and responsive IoT applications. As IoT plays an increasingly integral role in critical domains such as healthcare and industrial automation, optimizing resource allocation can directly impact service quality and reliability. Our research makes several significant contributions to the field of IoT resource allocation:

- Novel hybrid approach: We propose a novel resource allocation method that combines the power of ACO and TS algorithms. This hybrid approach harnesses the strengths of both algorithms to address the limitations of conventional resource allocation methods.
- QoS prioritization: Our method places a strong emphasis on QoS, aiming to improve energy efficiency,

reduce communication delays, and enhance overall system performance. This contributes to a more reliable and responsive IoT ecosystem.

- Efficiency and effectiveness: Through extensive testing and experimentation, we demonstrate the efficiency and effectiveness of our approach. Our results indicate significant improvements in QoS metrics when compared to traditional resource allocation methods.

The rest of the paper is organized in the following manner. Section II presents an introduction to IoT and resource allocation and describes existing resource allocation methods. Section III discussed the proposed method. Simulation results are reported in Section IV. Section V summarizes the main contributions of the study.

## II. BACKGROUND

### A. IoT Resource Allocation

Resource allocation is of utmost importance in the IoT ecosystem, where a multitude of interconnected devices collaborate to provide diverse services and applications. In the IoT, resources such as network bandwidth, computational power, storage capacity, and energy are scarce and must be allocated efficiently among numerous devices and applications [18]. Effective resource allocation in the IoT involves determining the optimal assignment of resources to meet the diverse requirements of IoT applications while considering factors such as QoS, energy efficiency, and network stability. A major challenge of IoT resource allocation stems from the dynamic and heterogeneous nature of IoT environments. IoT devices possess varying capabilities, communication protocols, and QoS requirements, further complicating the allocation process.

Additionally, IoT networks encounter fluctuations in resource availability due to device mobility, changing network conditions, and varying demands from different applications. Resource allocation algorithms in the IoT must be adaptive, scalable, and capable of handling network dynamics. Furthermore, considering resource limitations in IoT deployments, efficient allocation strategies are essential to prevent bottlenecks, ensure optimal resource utilization, and enhance the overall performance of IoT systems [19].

To address the resource allocation challenges in the IoT, a range of approaches have been proposed, each tailored to specific IoT scenarios and requirements. These approaches span from centralized algorithms to distributed mechanisms. Centralized resource allocation methods involve a central entity or server that receives requests from IoT devices and allocates resources based on predefined criteria or optimization objectives [20]. These algorithms centralize the decision-making process and effectively manage resource allocation in certain IoT environments. Distributed resource allocation algorithms aim to distribute the decision-making process among IoT devices themselves, enabling them to collaborate and negotiate for resources autonomously [21]. These algorithms promote self-organization and adaptability in resource allocation, making them suitable for dynamic and decentralized IoT systems. Optimization techniques such as genetic algorithms, swarm intelligence, and game theory have

been applied to solve the resource allocation problem in the IoT [22]. These techniques consider QoS, energy efficiency, load balancing, and fairness while allocating resources to IoT devices. They provide efficient and optimized resource allocation solutions based on mathematical models and optimization objectives. Machine learning and artificial intelligence-based approaches are gaining prominence in IoT resource allocation. These approaches leverage historical data and real-time analytics to make intelligent resource allocation decisions. By learning from past experiences and adapting to changing IoT conditions, machine learning-based algorithms can optimize resource allocation to meet dynamic IoT requirements. Edge computing offers a promising model for resource allocation in the IoT. By deploying computation and storage capabilities closer to IoT devices at the network edge, edge computing reduces latency, optimizes bandwidth usage, and enables localized resource allocation decisions. This decentralized approach minimizes the reliance on cloud resources and enhances the overall efficiency and responsiveness of IoT systems [23].

### B. Related Work

Wang, et al. [24] focused on addressing the distributed resource allocation problem in energy-efficient data forwarding for resource-constrained Industrial IoT (IIoT) systems [9]. They approached this problem by formulating it as a Decentralized, Partially Observable Markov Decision Process (Dec-POMDP), taking into account the decentralized and partially observable nature of the system. To tackle this challenge, they proposed an innovative algorithm named Dual-Attention assisted Deep Reinforcement Learning (DADR) for energy-efficient resource allocation. The DADR algorithm leverages a dual-attention assisted deep reinforcement learning (DRL) model within the Convolutional Attention Module, Dual-Attention, and Experience Reconstruction (CTDE) framework. The actor-network of the DADR algorithm incorporates a multi-scale convolutional attention module (CAM) to extract feature information from local states across various dimensions. Introducing a novel critic network, which employs a dual-attention module and an experience reconstruction module, enables comprehensive and precise evaluation of the system state from a global perspective. This critic network effectively addresses non-stationary and partially observable issues in multi-agent systems while maintaining scalability in dynamic environments without requiring modifications to the model structure. By combining CAM and Multi-Head Self-Attention (MHSA), the DADR algorithm enhances the representation learning capability of the DRL model.

Consequently, it provides improved optimization directions for energy efficiency and data transmission reliability. To assess the performance of the DADR algorithm, the researchers conducted simulations. The results of these simulations demonstrate the superiority of DADR over existing resource allocation algorithms and Multi-Agent Reinforcement Learning (MAREL) models in terms of network stability, transmission reliability, and network lifetime.

Kim and Ko [25] introduced a service resource allocation approach that aims to minimize data transmissions among users' mobile devices while effectively addressing the

constraints associated with such environments. To address the resource allocation problem, they transform it into a variant of the degree-constrained minimum spanning tree problem. Subsequently, they apply a genetic algorithm to efficiently generate near-optimal solutions within a shorter timeframe. The authors devise a fitness function and an encoding scheme specifically tailored to optimize the application of the genetic algorithm. Through the utilization of these components, the proposed approach demonstrates a remarkably high success rate, achieving near-optimal solutions in an average of 97% of cases. Moreover, it surpasses the brute force approach by significantly reducing the time required for solution generation.

In the study conducted by Tsai [26], the challenges associated with resource allocation in IoT systems are addressed. These systems are characterized by diverse user requirements, different types of appliances, limited network bandwidth, and computation power, all of which pose limitations to the performance of IoT systems and necessitate effective resource allocation solutions. To tackle this problem, the author proposes an algorithm that combines the concepts of data clustering and metaheuristics. The algorithm focuses on allocating the large-scale devices and gateways within the IoT system in a manner that minimizes the total communication cost between them. By optimizing the resource allocation, the algorithm aims to enhance the overall performance of the IoT system. The proposed algorithm is evaluated through simulations, and the results demonstrate its superiority over other resource allocation algorithms considered in the study. Specifically, the algorithm outperforms alternative approaches in terms of reducing the total data communication costs, highlighting its effectiveness in optimizing resource allocation for IoT systems.

Deng, et al. [27] address the challenge of trustworthiness management in edge computing (EC) systems, which play a crucial role in handling the increasing number of IoT devices connected to the edge of the network. They focus on ensuring compliance with service-level agreements (SLAs), which serve as an important indicator of trustworthiness for IoT services. To tackle this challenge, the authors propose a solution that involves modeling the state of the service provisioning system and the resource allocation scheme as a Markov decision process (MDP). They encode the trustworthiness gain, measured by the degree of SLA compliance, and use it as the objective for resource allocation adjustments. To obtain an optimal resource allocation policy, the authors employ reinforcement learning (RL) techniques. They train a policy using RL methods, which enables the dynamic generation of resource allocation schemes based on the system's current state. The trained policy is designed to maximize the trustworthiness gain of the services by allocating resources appropriately. The proposed approach is evaluated through experiments conducted on the YouTube request dataset. The results demonstrate that the edge service provisioning system utilizing the proposed approach outperforms baseline approaches by at least 21.72% in terms of performance.

Nematollahi, et al. [28] have proposed a novel architecture for offloading jobs and allocating resources for the IoT by incorporating Fog Computing (FC). They aim to address the limitations of low processing power and the need for efficient

data processing and management in IoT applications. The architecture consists of three main components: sensors, controllers, and FC servers. The authors introduce the concept of the subtask pool approach in the second layer, which enables the offloading of work from IoT devices to FC servers. To optimize resource allocation, they combine the Moth-Flame Optimization (MFO) algorithm with Opposition-based Learning (OBL), forming the OBLMFO algorithm. In the second layer, a stack cache approach is implemented to ensure resource allocation is balanced and prevent system load imbalance. The authors also leverage blockchain technology to guarantee the accuracy of transaction data, enhancing the reliability and transparency of resource distribution in the IoT system. To evaluate the performance of the OBLMFO model, the authors conducted experiments using the Python environment with a diverse set of jobs. The results demonstrate that the OBLMFO model achieved a 12.1% reduction in the delay factor and a 6.2% reduction in energy consumption compared to existing approaches.

Nguyen, et al. [29] propose a generalized federated learning (FL) algorithm to address the challenges encountered in FL, including non-independent and identically distributed data and heterogeneity of user equipment (UE). The objective of their approach is to reduce the global communication burden and enhance the convergence rate of FL. The proposed FL algorithm builds upon the current state-of-the-art federated averaging (FedAvg) by introducing a weight-based proximal term to the local loss function. This modification enables the

algorithm to perform stochastic gradient descent in parallel on a sampled subset of UEs during each global round, effectively reducing the communication overhead. The researchers provide a convergence upper bound that illustrates the tradeoff between the convergence rate and the number of global rounds. The analysis shows that convergence can still be ensured even with a small number of active UEs per round.

Liu [30] proposed a resource allocation algorithm for mobile edge computing. The algorithm aims to optimize base station performance over the long term by considering various factors, such as cable channel congestion, energy consumption, latency, communication cost, and task arrival characteristics. An energy consumption deficit queue based on Lyapunov drift penalties is introduced. This queue couples the energy consumption and time of small base stations, ensuring that energy consumption constraints are met during optimization. To calculate the offloading weight for task allocation, the authors employ game theory and propose an offloading weight formula derived from the Shapley value. The offloading weight is computed impartially, factoring in the return of various tasks. Simulations on the MATLAB platform were used to evaluate the proposed algorithm's performance. The algorithm can attain Nash equilibrium within a finite number of iterations, according to the results. Furthermore, the algorithm outperforms other comparison strategies in terms of the number of successfully offloaded tasks, time delay, and energy consumption.

TABLE I. IOT RESOURCE ALLOCATION METHODS

Method	Approach	Key features	Performance
DADR	Deep reinforcement learning model based on partially observable Markov decision processes	Centralized training and distributed execution framework Actor-network with a multi-scale convolutional attention module. Novel critic network based on dual-attention module	Outperforms existing resource allocation algorithms and multi-agent reinforcement learning models
Genetic Algorithm	Transformation of the resource allocation problem into a variant of the degree-constrained minimum spanning tree problem	Fitness function and encoding scheme for optimization	Achieves near-optimal solutions in 97% of cases on average and outperforms brute force approach
Data clustering and metaheuristics	Algorithm leveraging data clustering and metaheuristics	Minimization of total communication cost. Optimization of IoT system performance	Outperforms other resource allocation algorithms and reduces total data communication costs
Markov decision process with reinforcement learning	Modeling of state of service provisioning system and resource allocation as MDP	Trustworthiness gain as objective. Dynamic generation of resource allocation schemes	Outperforms baseline approaches by at least 21.72%
Fog Computing with moth-flame optimization and opposition-based learning	Architecture incorporating fog computing for resource allocation optimization	Subtask pool approach. Stack cache approach. Blockchain technology for accuracy of transaction data	Achieves reduction in delay factor and energy consumption
Generalized federated learning	Weight-based proximal term in local loss function and parallel stochastic gradient descent on a sampled subset of user equipment	Convergence rate improvement. Reduction of global communication burden	Requires less training time and energy consumption compared to full user participation
Task offloading and resource allocation algorithm	Optimization of long-term performance of small base stations and consideration of various factors	Energy consumption deficit queue. Offloading weight model based on Shapley value	Achieves Nash equilibrium and outperforms other comparison strategies

Table I presents a summary of IoT resource allocation methods along with their key features and performance characteristics. The reviewed methods often exhibit limitations that make them less suitable for the resource allocation problem. Some optimization techniques, such as genetic algorithm, can suffer from slow convergence, especially in large-scale IoT systems. This can hinder real-time decision-making, which is essential in many IoT applications. PSO and

traditional heuristic methods may produce suboptimal solutions when dealing with the complex, combinatorial nature of IoT resource allocation problems. Many existing methods do not inherently prioritize QoS metrics. They may not effectively address the specific QoS requirements of diverse IoT applications. We chose the proposed hybrid method, which combines ACO and TS algorithms, for several compelling reasons:

- **Combinatorial optimization:** IoT resource allocation is inherently a combinatorial optimization problem, as it involves allocating limited resources to tasks with varying requirements. ACO excels in solving such problems by mimicking the foraging behavior of ants and constructing solutions through probabilistic resource selection. This aligns with the nature of the resource allocation problem in IoT systems.
- **Local search enhancement:** While ACO provides global exploration, TS offers local search capabilities by maintaining a Tabu list. TS intensifies the search process by preventing revisits to previously explored solutions, which is crucial for avoiding suboptimal solutions in IoT resource allocation.
- **QoS focus:** Our primary objective is to prioritize QoS in IoT resource allocation. Existing methods may lack the necessary mechanisms to give due consideration to QoS metrics, such as energy efficiency and reduced communication delays. Our hybrid approach allows us to explicitly address these QoS concerns.

### III. PROPOSED METHOD

This section introduces a novel resource allocation algorithm called ACO-TS, which combines the ACO and TS algorithms. ACO-TS offers improved efficiency in terms of cost and execution time compared to existing approaches. We delve into the details of ACO-TS in the following subsections: Firstly, we define the problem and outline the objective function. Next, we provide a comprehensive explanation of the

proposed method, highlighting its key features and mechanisms.

#### A. Problem Definition

The distributed nature of resources and the need for efficient access make resource allocation in IoT challenging. The integration of multiple applications and the heterogeneity of connectivity further complicate resource allocation. Efficiency in an IoT system is measured by factors such as allocation interval, response time, and processing time, which are important QoS constraints. This study focuses on the resource allocation problem involving resource nodes and gateways. Resources are allocated to service instances, and gateways are responsible for connecting to these resources. Gateways serve as the interconnection points for IoT systems, managing the traffic of multiple resources. Effective resource allocation requires optimizing the distribution of resources among gateways to minimize communication costs. Additionally, the connectivity between gateways is crucial, and minimizing communication costs is a key objective. Various connection models, such as ring or bus connections, can be considered to achieve this goal. Fig. 1 illustrates an example of a resource-gateway connection. Communication costs depend on the chosen communication model. The objective of the problem is to determine the resource allocation pattern that minimizes communication costs. Another important aspect is load balancing, which involves distributing resources among gateways to prevent bottlenecks. The objective function section describes how load balancing and communication costs are calculated in the problem.

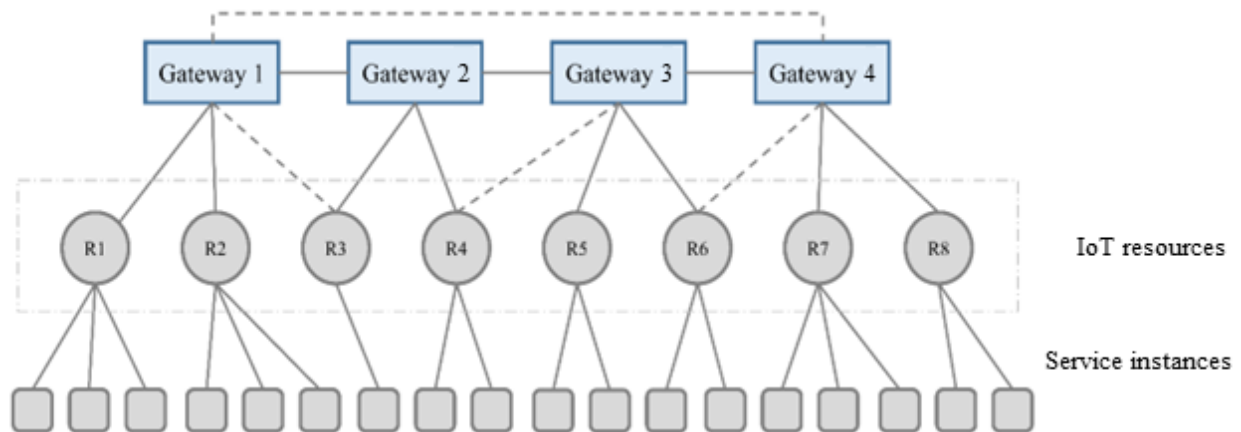


Fig. 1. An example of a resource-gateway connection.

#### B. Objective Function

In the considered network model, it is assumed that all resource nodes have the ability to communicate with each other. Therefore, to evaluate the different solutions to the resource allocation problem, it is necessary to calculate the total cost of network communications. It is assumed that all resources send messages to each other, and the objective function aims to minimize the total cost of these messages. The total cost is calculated using Eq. 1, which represents the mathematical expression for determining the cost. The proposed algorithm focuses on minimizing this objective function as its main goal.

$$T_c = \frac{\sum_{j=1}^{|V_g|} (d_j^r \times d^g)}{p} \quad (1)$$

In Eq. (1), the variable  $d^g$  represents the overall communication cost among gateways,  $d_j^r$  represents the total data transfer cost between the  $j^{th}$  gateway and all resources connected, and  $V_g$  represents the total number of gateways in the network. It is important to note that communication within a gateway can also result in messages being exchanged between gateways. The exponential nature of the communication cost within a gateway greatly affects the total communication cost. Thus, the equation  $(d_j^r \times d^g)$  is the most

appropriate way to calculate the maximum communication costs. The objective is to minimize the value of the objective function. Since gateways have the ability to send messages to all resources, we need to multiply  $d_j^r$  by  $d^e$ . By summing up these values for all gateways, we can obtain an estimation of the total communication costs for the gateways. The numerator part of the  $T_c$  fraction in the objective function can be calculated by performing this calculation for each gateway.

### C. Proposed Hybrid Algorithm

The proposed hybrid method combines the ACO algorithm and the TS technique to tackle the IoT resource allocation problem, considering QoS requirements. This combination leverages the strengths of both algorithms to achieve efficient and effective resource allocation. The ACO algorithm, which mimics the foraging behavior of ants, simulates the behavior of ants in search of food [31]. The algorithm constructs solutions by probabilistically selecting resources based on pheromone trails and heuristics. The pheromone trail represents the quality of each resource allocation solution, and ants deposit and update pheromone values as they move through the solution space. This behavior encourages the exploration of promising solutions and exploits the accumulated knowledge of previous solutions. By utilizing the ACO algorithm, the proposed method can efficiently explore the large solution space and identify potential resource allocation configurations that satisfy the QoS requirements. To further enhance the search process and overcome local optima, the TS technique is integrated into the proposed method. TS is a local search-based metaheuristic that intensifies the search by maintaining a tabu list [32]. The tabu list keeps track of recently visited solutions, preventing them from being revisited in subsequent iterations. This mechanism encourages diversification and exploration of new regions in the solution space, allowing the algorithm to escape local optima and search for better resource allocation solutions. Neighborhood search and diversification strategies are also applied within the TS framework to further improve the search process.

The proposed hybrid method balances exploration and exploitation by fusing the ACO algorithm and TS technique. The ACO algorithm explores the solution space using pheromone trails and heuristics, while the TS technique intensifies the search process by leveraging the tabu list and neighborhood search strategies. This combination enables the algorithm to efficiently converge towards optimal or near-optimal solutions that satisfy the QoS requirements of IoT applications. The ACO algorithm updates the pheromone trails based on the quality of each resource allocation solution. The pheromone update formula is given by:

$$\tau_{ij} = (1 - \rho)\tau_{ij}^{old} + \sum_{k=1}^m \Delta\tau_{ij}^k \quad (2)$$

where  $\tau_{ij}$  represents the pheromone level on edge  $(i, j)$  in the resource allocation graph,  $\rho$  is the evaporation rate that controls the decay of pheromone over time, and  $\Delta\tau_{ij}^k$  represents the pheromone increment for ant  $k$  on edge  $(i, j)$  in the solution construction phase. The ACO algorithm constructs solutions by probabilistically selecting resources based on the pheromone trails and heuristics. The probability of selecting resource  $j$  for ant  $k$  at node  $i$  is calculated using the following equation:

$$P_{ij}^k(t) = \begin{cases} \frac{[\tau_{ij}(t)]^\alpha [\eta_{ij}]^\beta}{\sum_{l \in N_k} [\tau_{il}(t)]^\alpha [\eta_{il}]^\beta} & \text{if } j \in N_k \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

Where  $\alpha$  and  $\beta$  control the relative importance of pheromone trails and heuristics,  $\eta_{ij}$  represents the heuristic information, which captures the desirability of allocating resource  $j$  at node  $i$  based on factors such as energy consumption, network capacity, and QoS requirements. The TS algorithm maintains a tabu list to prevent revisiting previously visited solutions. The tabu list is updated using the following equation:

$$TL = TL_U(i, j) \quad (4)$$

Where  $TL$  represents the tabu list,  $(i, j)$  represents the resource allocation move (e.g., allocating resource  $j$  at node  $i$ ) that is added to the tabu list. The pseudocode for the proposed hybrid method is shown in Algorithm 1. The proposed algorithm begins by initializing the necessary variables and data structures, including the resource allocation graph, QoS requirements, and algorithm parameters. It then enters a loop that repeats for a specified number of iterations. Within each iteration, a set of ants is created to construct resource allocation solutions. Each ant starts at a random node and selects resources probabilistically according to the pheromone trails and heuristics. The constructed solutions are evaluated against the QoS requirements, and pheromone trails are adjusted accordingly. After the ACO phase, the algorithm proceeds to the TS initialization. In this case, the best solution is set as the current solution, and the tabu list is initialized as empty. The TS loop begins, continuing until a stopping criterion is met. In each iteration of the TS loop, neighboring solutions are generated by making resource allocation moves from the current solution, taking into account the tabu list restrictions. The best non-tabu solution is selected, and if it improves upon the current best solution, it becomes the new best solution. The corresponding resource allocation move is added to the tabu list, and the selected solution becomes the current solution for the next iteration. The algorithm repeats this process for the specified number of iterations, continuously improving the resource allocation solutions. Finally, the best resource allocation solution found is output as the final result.

#### Algorithm 1. Pseudocode of the proposed algorithm

<p><b>Initialize:</b></p> <ul style="list-style-type: none"> <li>- Initialize pheromone levels <math>\tau_{ij}</math> for all edges <math>(i, j)</math> in the resource allocation graph</li> <li>- Initialize an empty tabu list <math>TL</math></li> </ul> <p><b>Repeat</b> for a specified number of iterations:</p> <ul style="list-style-type: none"> <li>Create a set of ants:</li> <li>- For each ant <math>k</math>: <ul style="list-style-type: none"> <li>- Start at a random node <math>i</math></li> <li>- Initialize an empty resource allocation solution <math>S_k</math></li> </ul> </li> </ul> <p><b>Construct solutions:</b></p> <ul style="list-style-type: none"> <li>- For each ant <math>k</math>: <ul style="list-style-type: none"> <li>- While not all nodes are visited: <ul style="list-style-type: none"> <li>- Calculate the selection probability <math>P_{ij}^k</math> for unvisited neighboring nodes</li> <li>- Select a resource <math>j</math> based on the selection probability</li> <li>- Add resource <math>j</math> to the solution <math>S_k</math></li> <li>- Update the visited and unvisited node lists</li> </ul> </li> </ul> </li> </ul>
--



**Update pheromone trails:**

- For each ant  $k$ :
  - Calculate the solution quality metric  $Q_k$  based on QoS requirements
  - Update the pheromone trails:
    - For each edge  $(i, j)$  in the solution  $S_k$ :
      - Calculate the pheromone increment  $\Delta\tau_{ij}^k$
      - Update pheromone level  $\tau_{ij}$  using the pheromone update formula

**Evaporate pheromone trails:**

- For each edge  $(i, j)$  in the resource allocation graph:
  - Evaporate pheromone level  $\tau_{ij}$  using the evaporation rate  $\rho$

**Tabu Search:**

- Initialize the best solution  $S_{best}$  with the best solution found so far
- Set the current solution  $S_{curr}$  as  $S_{best}$
- Set the tabu list  $TL$  as empty
- Repeat until a stopping criterion is met:
  - Generate a set of neighboring solutions by making resource allocation moves on  $S_{curr}$ 
    - Select the best non-tabu solution  $S_{next}$  from the neighboring solutions
    - If  $S_{next}$  is better than  $S_{best}$ , update  $S_{best}$
    - Add the resource allocation move corresponding to  $S_{next}$  to the tabu list  $TL$
    - Set  $S_{curr}$  as  $S_{next}$

The overall time complexity of the combined ACO and TS algorithms can be expressed as a combination of the complexities of both algorithms. Assume the number of ants as  $N_a$ , the number of iterations as  $N_i$ , the size of the solution space as  $N_s$ , the number of resources available for allocation as  $M$ , the size of the neighborhood as  $N_n$ , and the diversification parameter as  $N_d$ . The time complexity of the ACO algorithm can be approximated as  $O(N_a * N_i * f(N_s))$ , where  $f(N_s)$  represents the complexity of constructing solutions. The ACO algorithm updates the pheromone trails based on the quality of each resource allocation solution, which has a complexity of  $O(M)$  since the pheromone values for each resource need to be updated. The time complexity of the TS algorithm primarily depends on the neighborhood search operation and the diversification strategies employed. Assuming the complexity of the neighborhood search operation is  $O(g(N_n))$  and the complexity of diversification strategies is  $O(h(N_d))$ , the overall time complexity of the TS algorithm can be approximated as  $O(N_i * (g(N_n) + h(N_d)))$ . Consequently, the time complexity of the combined ACO and TS algorithm can be represented as:

$$O(N_a \times (N_i \times O(f(N_s)) + M) + O(g(N_n)) + O(h(N_d))) \quad (5)$$

#### IV. RESULTS AND DISCUSSION

In this section, we present the simulation and evaluation of the suggested algorithm. The simulations were performed using MATLAB software on a desktop computer with a core i5 CPU and 4GB of RAM. The performance of the proposed resource allocation technique is assessed by comparing it with previous methods in two scenarios. The first scenario involves four experiments with varying numbers of gateway and source nodes. The details of these experiments are summarized in Table II. Each experiment is characterized by the number of gateway and resource nodes. These parameters allow for the evaluation and comparison of the proposed method under

different network configurations. The overall fitness of each mechanism is depicted in Fig. 2 to 5. From these figures, it can be observed that the proposed method consistently outperforms genetic, PSO, and ACO algorithms. Scalability is another important aspect considered in the proposed technique. Fig. 3 proves that as the network size increases, the convergence diagram of the proposed technique (blue line) diverges from the convergence diagrams of the ACO and genetic algorithms (red and black lines), indicating its scalability. The optimality of the convergence graph is determined by the communication costs in the network. The fitness function used in the evaluation takes into account execution time and energy consumption, which are crucial factors in determining the sub-function of fitness.

TABLE II. SIMULATION PARAMETER VALUES

Parameters	Experiment 1	Experiment 2	Experiment 3	Experiment 4
Number of gateway nodes	10	25	60	120
Number of resource nodes	30	100	500	800

We compare the proposed method with ACO and genetic algorithms in the second scenario. Similar to the first scenario, four experiments are conducted, following the setup described in Table II. Fig. 6 to 9 illustrate the performance of the algorithms in these experiments. From the figures, it can be observed that the genetic algorithm exhibits the highest execution time among the three algorithms. This prolonged decision-making process can lead to traffic congestion on the server, causing delays and inefficiencies. On the other hand, the ACO algorithm performs better than the genetic algorithm, as it takes less time to select the best resource. In comparison to existing benchmark algorithms, our proposed method demonstrates a faster decision-making process in selecting the best resource. This reduced time is beneficial in achieving efficient resource allocation and improving overall system performance.

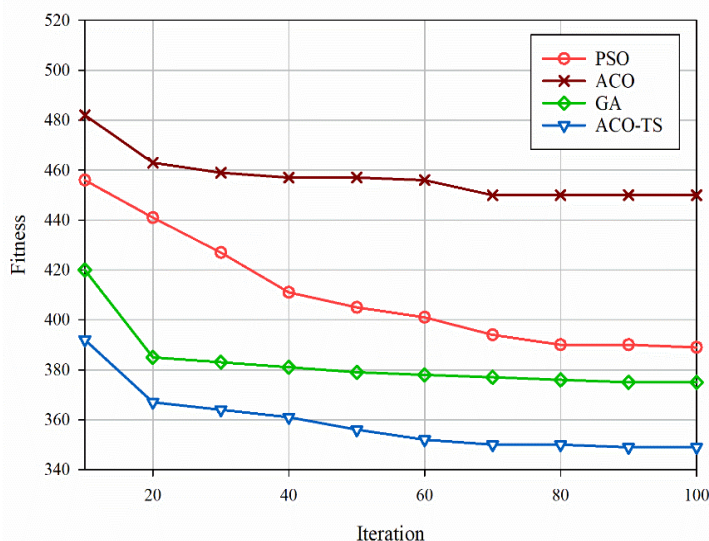


Fig. 2. Fitness comparison (First experiment).

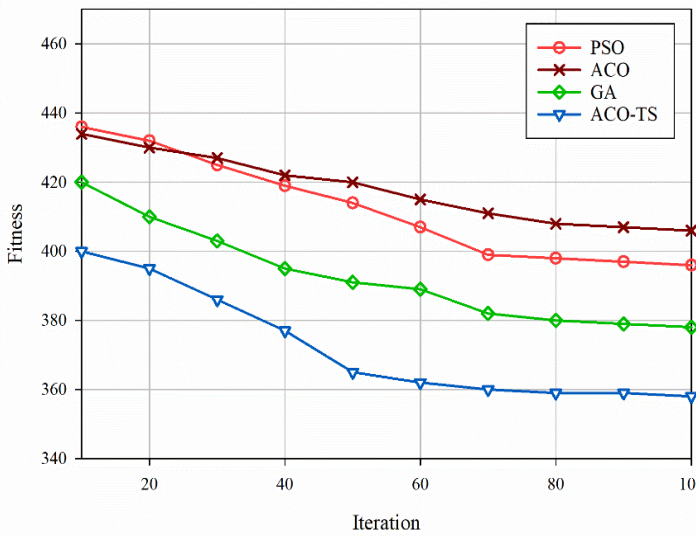


Fig. 3. Fitness comparison (Second experiment).

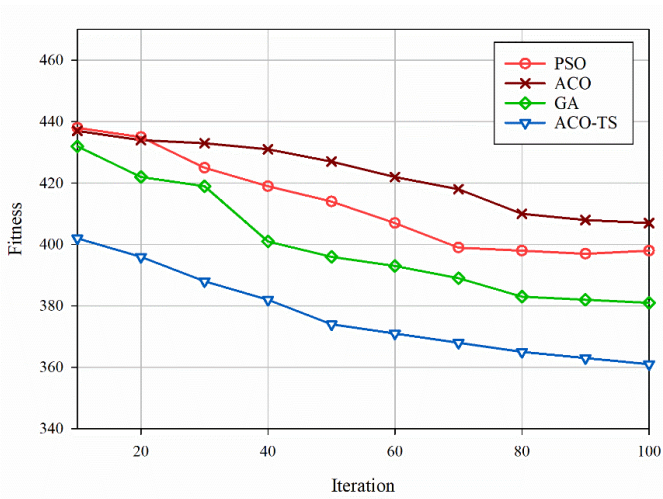


Fig. 4. Fitness comparison (Third experiment).

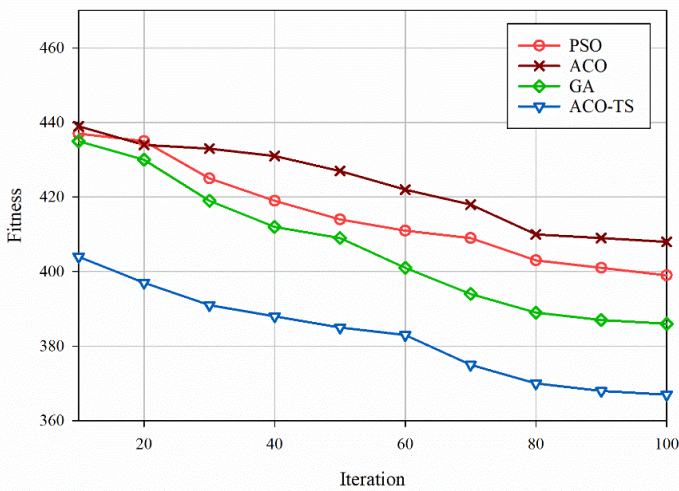


Fig. 5. Fitness comparison (Fourth experiment)

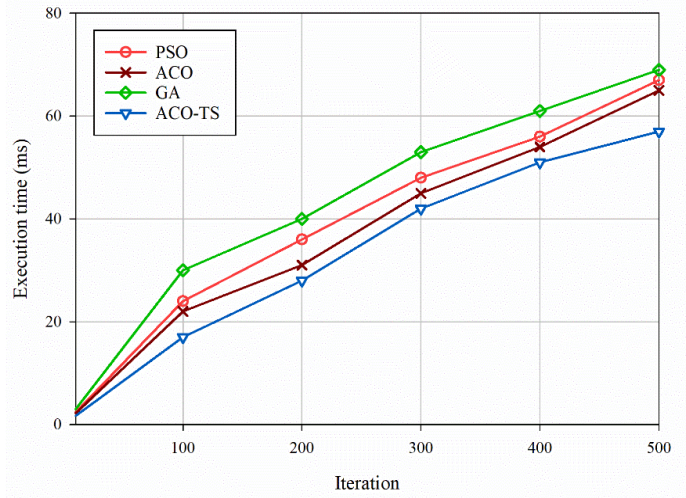


Fig. 6. Execution time comparison (First experiment).

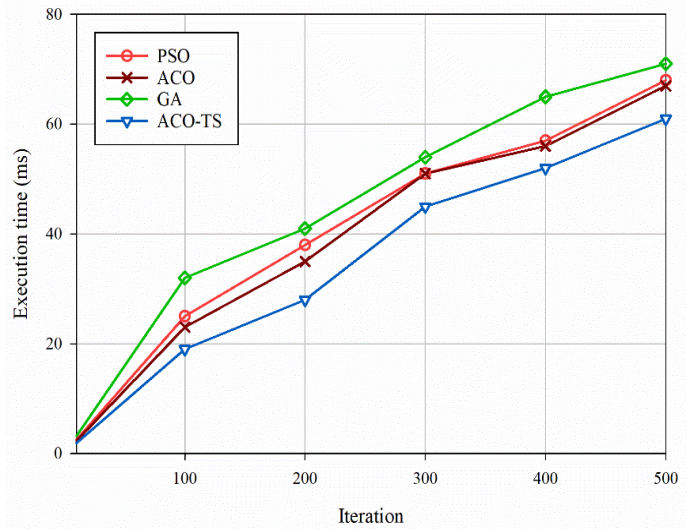


Fig. 7. Execution time comparison (Second experiment).

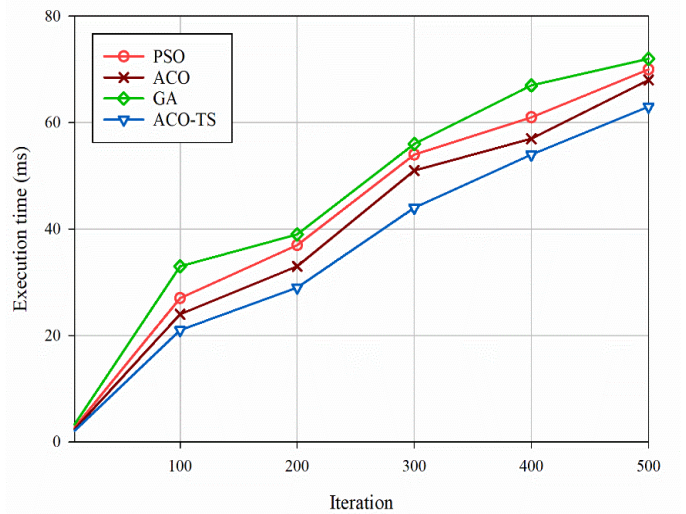


Fig. 8. Execution time comparison (Third experiment)

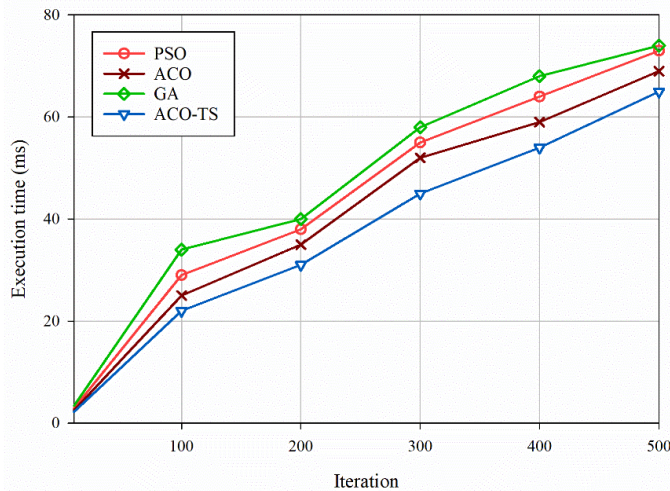


Fig. 9. Execution time comparison (Fourth experiment)

In our evaluation of the proposed algorithms, we utilized multiple datasets that reflect diverse IoT scenarios and characteristics. These datasets differ in terms of size, complexity, and the specific application domains they represent. Such variations inherently contribute to differences in algorithm performance. For instance, in datasets characterized by large-scale IoT networks with numerous interconnected devices, the proposed hybrid algorithm may excel in optimizing resource allocation by leveraging ACO's global exploration capabilities. On the other hand, in smaller-scale networks or datasets representing resource-intensive IoT applications, the local search enhancements offered by the TS algorithm might yield more pronounced benefits. Moreover, the nuances of each dataset, including the nature of IoT devices, their energy constraints, and communication patterns, can significantly impact the suitability of the proposed algorithms. The variations observed in comparative results across these datasets suggest that our algorithms exhibit adaptability to different IoT contexts. These findings open avenues for future research in fine-tuning the proposed algorithms to specific IoT scenarios, as well as exploring adaptive resource allocation strategies that can dynamically adjust to the unique requirements of diverse IoT data types and application domains. In essence, the dataset variations shed light on the algorithmic flexibility of our approach, indicating its potential to cater to the evolving and multifaceted landscape of IoT resource allocation.

## V. CONCLUSION

This paper suggested a novel QoS-aware resource allocation method for IoT systems using a hybrid approach of the ACO and TS algorithms, called ACO-TS. It efficiently allocates limited resources in IoT systems in accordance with the QoS requirements of individual devices. Through the integration of the ACO algorithm and TS technique, we have demonstrated the effectiveness and efficiency of our approach in optimizing resource allocation decisions. The ACO algorithm leverages the behavior of ant colonies to explore the solution space, while the TS technique intensifies the search process to overcome local optima. By combining these two techniques, we achieve a balance between exploration and

exploitation, resulting in improved convergence speed and solution quality. Our experimental evaluations in realistic IoT scenarios have suggested the merits of the ACO-TS approach. In comparison with existing resource allocation methods, ACO-TS achieves significant improvements in energy consumption reduction, network capacity maximization, and satisfaction of QoS requirements.

## REFERENCES

- [1] 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.
- [2] F. Kamalov, B. Pourghebleh, M. Gheisari, Y. Liu, and S. Moussa, "Internet of Medical Things Privacy and Security: Challenges, Solutions, and Future Trends from a New Perspective," *Sustainability*, vol. 15, no. 4, p. 3317, 2023.
- [3] T. Taami, S. Azizi, and R. Yarinezhad, "Unequal sized cells based on cross shapes for data collection in green Internet of Things (IoT) networks," *Wireless Networks*, pp. 1-18, 2023.
- [4] B. Pourghebleh and V. Hayyolalam, "A comprehensive and systematic review of the load balancing mechanisms in the Internet of Things," *Cluster Computing*, pp. 1-21, 2019.
- [5] A. Thakur and M. S. Goraya, "RAFL: A hybrid metaheuristic based resource allocation framework for load balancing in cloud computing environment," *Simulation Modelling Practice and Theory*, vol. 116, p. 102485, 2022.
- [6] 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, doi: <https://doi.org/10.1016/j.suscom.2022.100822>.
- [7] V. Srinadh and P. N. Rao, "Implementation of Dynamic Resource Allocation using Adaptive Fuzzy Multi-Objective Genetic Algorithm for IoT based Cloud System," in *2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 2022: IEEE, pp. 111-118.
- [8] A. K. Sangaiah, A. A. R. Hosseinabadi, M. B. Shareh, S. Y. Bozorgi Rad, A. Zolfagharian, and N. Chilamkurti, "IoT resource allocation and optimization based on heuristic algorithm," *Sensors*, vol. 20, no. 2, p. 539, 2020.
- [9] M. Mohseni, F. Amirghafouri, and B. Pourghebleh, "CEDAR: A cluster-based energy-aware data aggregation routing protocol in the internet of things using capuchin search algorithm and fuzzy logic," *Peer-to-Peer Networking and Applications*, pp. 1-21, 2022.
- [10] S. Li, Q. Ni, Y. Sun, G. Min, and S. Al-Rubaye, "Energy-efficient resource allocation for industrial cyber-physical IoT systems in 5G era," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 6, pp. 2618-2628, 2018.
- [11] 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.
- [12] B. M. Jafari, M. Zhao, and A. Jafari, "Rumi: An Intelligent Agent Enhancing Learning Management Systems Using Machine Learning Techniques," *Journal of Software Engineering and Applications*, vol. 15, no. 9, pp. 325-343, 2022.
- [13] C. Han and X. Fu, "Challenge and Opportunity: Deep Learning-Based Stock Price Prediction by Using Bi-Directional LSTM Model," *Frontiers in Business, Economics and Management*, vol. 8, no. 2, pp. 51-54, 2023.
- [14] 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.
- [15] 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.

- [16] H. Kosarirad, M. Ghasempour Nejadi, A. Saffari, M. Khishe, and M. Mohammadi, "Feature Selection and Training Multilayer Perceptron Neural Networks Using Grasshopper Optimization Algorithm for Design Optimal Classifier of Big Data Sonar," *Journal of Sensors*, vol. 2022, 2022.
- [17] S. Saeidi, S. Enjedani, E. Alvandi Behineh, K. Tehranian, and S. Jazayerifar, "Factors Affecting Public Transportation Use during Pandemic: An Integrated Approach of Technology Acceptance Model and Theory of Planned Behavior," *Tehnički glasnik*, vol. 18, pp. 1-12, 09/01 2023, doi: 10.31803/tg-20230601145322.
- [18] P. Zuo, G. Sun, Z. Li, C. Guo, S. Li, and Z. Wei, "Towards Secure Transmission in Fog Internet of Things Using Intelligent Resource Allocation," *IEEE Sensors Journal*, 2023.
- [19] B. Pourghebleh, V. Hayyolalam, and A. A. Anvigh, "Service discovery in the Internet of Things: review of current trends and research challenges," *Wireless Networks*, vol. 26, no. 7, pp. 5371-5391, 2020.
- [20] M. A. Raza, M. Abolhasan, J. Lipman, N. Shariati, W. Ni, and A. Jamalipour, "Statistical Learning-based Adaptive Network Access for the Industrial Internet-of-Things," *IEEE Internet of Things Journal*, 2023.
- [21] X. Liu and X. Zhang, "NOMA-based resource allocation for cluster-based cognitive industrial internet of things," *IEEE transactions on industrial informatics*, vol. 16, no. 8, pp. 5379-5388, 2019.
- [22] B. Pourghebleh, K. Wakil, and N. J. Navimipour, "A comprehensive study on the trust management techniques in the Internet of Things," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 9326-9337, 2019.
- [23] B. Pourghebleh, N. Hekmati, Z. Davoudnia, and M. Sadeghi, "A roadmap towards energy - efficient data fusion methods in the Internet of Things," *Concurrency and Computation: Practice and Experience*, p. e6959, 2022.
- [24] Y. Wang, F. Shang, J. Lei, X. Zhu, H. Qin, and J. Wen, "Dual-attention assisted deep reinforcement learning algorithm for energy-efficient resource allocation in Industrial Internet of Things," *Future Generation Computer Systems*, vol. 142, pp. 150-164, 2023.
- [25] M. Kim and I.-Y. Ko, "An efficient resource allocation approach based on a genetic algorithm for composite services in IoT environments," in *2015 IEEE international conference on web services*, 2015: IEEE, pp. 543-550.
- [26] C.-W. Tsai, "SEIRA: An effective algorithm for IoT resource allocation problem," *Computer Communications*, vol. 119, pp. 156-166, 2018.
- [27] S. Deng et al., "Dynamical resource allocation in edge for trustable Internet-of-Things systems: A reinforcement learning method," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 9, pp. 6103-6113, 2020.
- [28] M. Nematollahi, A. Ghaffari, and A. Mirzaei, "Task and resource allocation in the internet of things based on an improved version of the moth-flame optimization algorithm," *Cluster Computing*, pp. 1-23, 2023.
- [29] V.-D. Nguyen, S. K. Sharma, T. X. Vu, S. Chatzinotas, and B. Ottersten, "Efficient federated learning algorithm for resource allocation in wireless IoT networks," *IEEE Internet of Things Journal*, vol. 8, no. 5, pp. 3394-3409, 2020.
- [30] J. Liu, "Task offloading and resource allocation algorithm based on mobile edge computing in Internet of Things environment," *The Journal of Engineering*, vol. 2021, no. 9, pp. 500-509, 2021.
- [31] M. Dorigo, M. Birattari, and T. Stutzle, "Ant colony optimization," *IEEE computational intelligence magazine*, vol. 1, no. 4, pp. 28-39, 2006.
- [32] W. Fan and R. B. Machemehl, "Tabu search strategies for the public transportation network optimizations with variable transit demand," *Computer - Aided Civil and Infrastructure Engineering*, vol. 23, no. 7, pp. 502-520, 2008.

# Artificial Rabbits Optimizer with Deep Learning Model for Blockchain-Assisted Secure Smart Healthcare System

Mousa Mohammed Khubrani

College of Computer Science and IT, Jazan University, Jazan, Saudi Arabia

**Abstract**—Smart healthcare is based on the electronic health and medical histories of residents, combined with information technology (IT) which can be used to construct a variety of systems including humanised health management systems and convenient medical service systems. The transparency, traceability, decentralization and security of BC technology and machine learning (ML) will enable the medical sector to upgrade and optimise different forms of quality and service. Therefore, this study introduces an artificial rabbit optimizer with deep learning for Blockchain Assisted Secure Smart Healthcare System (ARODL-BSSHS) technique. The presented ARODL-BSSHS technique designs a new healthcare monitoring technique by using blockchain (BC) technology and classifies the presence of malicious activities in the healthcare system, and takes needed actions to predict the disease. For intrusion detection, the ARODL-BSSHS technique exploits the ARO algorithm with Hop field neural network (IHNN) model. On the other hand, the ARODL-BSSHS technique applies a deep extreme learning machine (DELIM) model for disease detection purposes. Finally, the heap-based optimization (HBO) technique is exploited as a hyperparameter optimizer for the DELIM model. The ARODL-BSSHS technique involves BC technology for the secure transmission of healthcare data. A series of simulations were carried out on benchmark datasets: heart disease and NSL-KDD database for examining the performance of the ARODL-BSSHS technique. The experimental values highlighted that the ARODL-BSSHS method obtains superior performance than other approaches.

**Keywords**—Blockchain; smart healthcare; artificial rabbit's optimizer; deep learning; intrusion detection

## I. INTRODUCTION

The connection of clinically related technologies will have a major impact on healthcare professionals and patients [1]. Along with the diversified nature and fast growth of the health care atmosphere, protection becomes a major problem as advanced security problems develop and earlier security problems become more acute. Data protection can be defined as the capability to transmit and store data without enabling unauthorized access to make sure confidentiality, data consistency, legality, and legitimacy [2]. Only authorized users have access to the protected data. Due to their unauthorized access and unauthenticated users, cybercrime develops and often affects healthcare sensors and systems [3]. A considerable amount of healthcare data is distributed, collected and gathered among various health care sectors. The data transmission must take place in a protected manner. The

number of cyber-attacks is increasing drastically due to enormous data transformation. It denotes the demand for a reliable system for protecting health care datasets [4]. Potential mining methods are demanded to inspect clinical data to assist in enhancing patient care, disease discovery, and offering medical treatment [5]. ML can be a complex computational method that was employed in various fields like health care, image recognition, and language processing [6]. Still, ML methods obtain a higher level of accuracy with a large volume of the training set that can be vital in health care, where accuracy may, in some cases, denote the difference between losing and saving the life of the patient. In many cases, centralized training methods acquire a large quantity of data from robust cloud servers that result in major consumer privacy violations, particularly in the clinical domain [7]. As an accountable and open data protection system, the progression of the BC technology opens the way for novel ways to solve the main problems of ethics, privacy, and security in domains that require privacy, anonymity and security of records including health care system [8] [9]. But BC has attained remarkable achievement for different smart healthcare technologies like patient record access control, data distribution, etc. [10].

Today, BC and ML technology are preferred [11]. The security, traceability, transparency, and decentralization of these two technologies will assist the healthcare sector to upgrade and optimize in several aspects [12]. The implementation of and making the functioning of the health care sector more efficient [13]. Few studies have explored the implementation of ML and BC. For instance, a health management platform based on BC can allow users to track personal data securely, and smart contracts are utilized in clinical detection to automatically manage emergencies [14].

This research is driven by the urgent need to address the growing vulnerabilities in healthcare systems due to escalating cyber threats and the rapid advancements in healthcare technologies. This study introduces an Artificial Rabbit Optimizer with Deep Learning for Blockchain-Assisted Secure Smart Healthcare System (ARODL-BSSHS) technique. This novel technique focuses on creating a secure and intelligent healthcare system, specializing in intrusion detection and disease diagnosis. It strategically employs the Artificial Rabbit Optimizer (ARO) algorithm and the Hopfield Neural Network (HNN) model for intrusion detection, and a Deep Extreme Learning Machine (DELIM) model for accurate disease detection, with Blockchain technology incorporated to secure

data transmission. The efficacy and improved performance of the ARODL-BSSHS technique have been validated through extensive experiments on recognized datasets, showcasing its potential for real-world applications in enhancing healthcare security and efficiency.

## II. LITERATURE REVIEW

In [15], the authors introduced a smart BC Manager (BM) depends on the DRL for optimizing the BC behavior of the network in real-time while concerning clinical data needs, like security levels and urgency. Utilizing 3 RL-related methods like Dueling Double Deep Q-Network (D3QN), Double Deep Q-Networks (DDQN), and DQN, the optimization approach can be developed as a Markov Decision Process (MDP). Lakhan et al. [16] present a DRLBTS abbreviated as DRL-aware BC-related task scheduling structure with various goals. The presented method offers security and makespan potential scheduling for medicinal purposes. Singh et al. [17] modelled a DL-related IoT-based structure for the secured smart city where BC offers a dispersed atmosphere at the transmission stage of CPS, and SDN established the protocol for transporting data. A DL-related cloud was applied at the application layer of the presented structure to solve scalability, centralization, and communication latency.

Mantey et al. [18] presented a BC privacy system (BPS) as DL for diet recommendation mechanisms for patients. This study applied DL and ML approaches like MLP, RNN, LR etc., to the Internet of Medical Things (IoMT) data obtained. The product section contains a collection of eight attributes. The IoMT dataset features are examined with BPS and encoded in advance to the implementation of DL and ML-related structures. In [19], presented a BC-orchestrated DL method (BDSDT) for Secured Data transmission in IoT-based healthcare systems. First, a new scalable BC structure is devised to ensure secure data transmission and data integrity using Zero Knowledge Proof (ZKP) system. Afterwards, BDSDT integrated with the off-chain storage IPFS abbreviated as InterPlanetary File mechanism, to solve problems with data storing costs to solve data security problems. Sammeta and Parthiban [20] developed a new method HBESDM-DLD abbreviated as hyperledger BC-based secure clinical data management with DL-related diagnosis method. This method includes different stages of operations such as hyperledger BC-based secure data management, encryption, diagnosis and optimal key generation. For encryption, SIMON block cipher method can be implemented. For optimal key generation, a group teaching optimization algorithm (GTOA) was adopted.

In [21], the authors proposed a Decentralized Interoperable Trust framework (DIT) based on BC for the Internet of Things (IoT) platform. The DIT IoHT employs a private BC ripple chain to establish secure and reliable data transmission by authenticating nodes in relation to their interoperable structures. Purbey, Khandelwal, and Choudhary in [22] introduced a method for secure and efficient ontology generation using BC, named BOGMAS. This approach employs a semi-supervised technique to generate ontologies from structured or unstructured datasets. It combines techniques such as extra trees (ET) stratification and linear support vector machine (LSVM) for predicting variances.

Almaiah et al. [23] proposed a Deep Learning (DL) architecture integrated with BC to ensure dual levels of privacy and security. Firstly, they establish a BC model where participating entities undergo registration, validation, and verification through smart contracts using Proof of Work. Subsequently, they model BiLSTM for intrusion detection and apply a DL method incorporating a Variational Autoencoder (VAE) technique for privacy preservation.

The reviewed studies, despite their innovative contributions to blockchain and deep learning in healthcare, exhibit several overarching limitations. Many face issues related to scalability, adaptability, and specificity, which can restrict their applicability across diverse healthcare environments and requirements. Several solutions also struggle with the balance between complexity and user-friendly implementation, posing challenges in deployment and interpretation. Additionally, the methods proposed often focus narrowly on specific aspects of healthcare or technology, neglecting a holistic approach that addresses the multifaceted nature of healthcare systems, thus necessitating further holistic and integrative research endeavors.

## III. PROPOSED MODEL

In this study, the ARODL-BSSHS technique has been developed to accomplish security in the healthcare system. The presented ARODL-BSSHS technique involves the design of secured and smart healthcare system using two major processes, namely intrusion detection and disease diagnosis. To accomplish this, the ARODL-BSSHS technique follows a series of processes: HNN based intrusion detection, ARO based parameter tuning, DELM-based disease detection, and HBO based parameter optimization. Fig. 1 illustrates the workflow of ARODL-BSSHS algorithm.

### A. BC Technology

In this work, the ARODL-BSSHS technique involves BC technology for secure transmission of healthcare data. Electronic Health Records (EHR) are well functioning on smart contracts [24]. It developed the framework for a decentralized medical service stage and aids as an interface to the patient records that can be shared by suppliers and patients. BC is separated into research-centric and patient-centric BC network classes, as stated by "BC Technology in Healthcare". The security concern regarding EHR is tackled by the patient-centric BC network, which gives the authority over sharing medicinal data with multiple users. BC technology can be able to modernize healthcare management by permitting unambiguous and transparent data access through every stakeholder involved, comprising hospitals, therapists, medical experts, and general practitioners [25]. In such cases, several medicinal stakeholders do not necessarily use resource- and time-consuming information and verification progressions.

Furthermore, this approach could contribute to the early detection of health-related issues, thereby reducing instances of medical malpractice arising from coordination issues [26]. It instills trust in individuals regarding their comprehensive care, as the integrity of healthcare records from previous visits to different medical practitioners remains intact within the network. Additionally, BC serves as a valuable tool for

constructing patient-centric networks through various means, such as augmenting data availability, thereby enhancing data liquidity, establishing unique patient identifiers, and implementing digital access control. The Hyperledger Fabric system can be leveraged to establish a permissioned BC network. Within this system, two distinct types of peers exist: validating peers, responsible for ledger management, consensus procedures, and transaction validation. The data is stored within a distributed system, facilitating the upload of patient medical histories, verification of healthcare records, and the facilitation of data access requests and permissions.

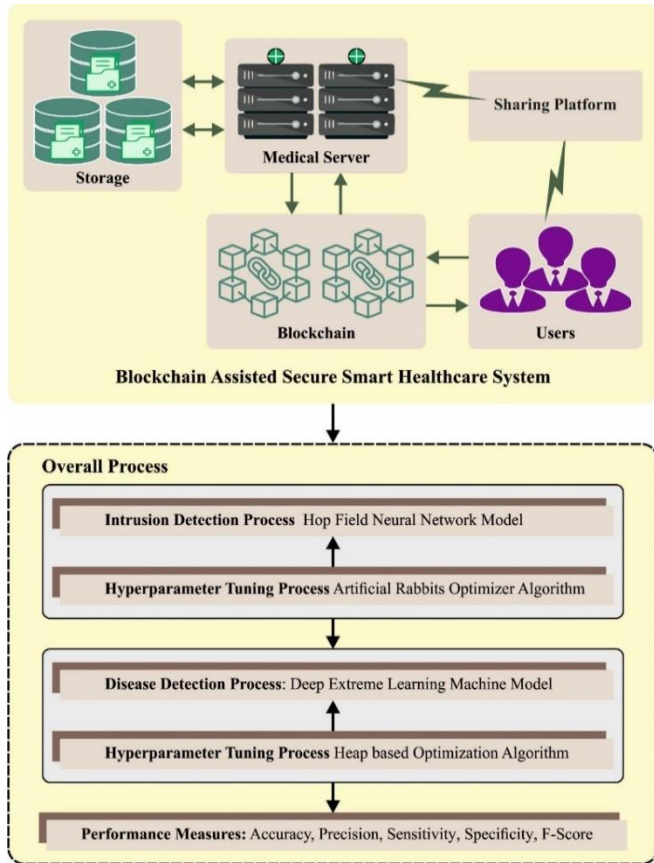


Fig. 1. Workflow of ARODL-BSSHS approach.

### B. Intrusion Detection using Optimal HNN Model

For intrusion detection process, the HNN classifier is used. The HNN exhibits abundant dynamical behavior owing to its hyperbolic tangent function and special network structure [27]. The HNN with  $n$  neurons is defined by the series of dimensionless non-linear ordinary differential equations as in the following:

$$\dot{x} = -x + Y \tanh(x) + I \quad (1)$$

Where

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_i \\ \vdots \\ x_n \end{bmatrix}, I = \begin{bmatrix} I_1 \\ I_2 \\ \vdots \\ I_i \\ \vdots \\ I_n \end{bmatrix}$$

$$Y = \begin{bmatrix} y_{11} & y_{12} & \dots & y_{1j} & \dots & y_{1n} \\ y_{21} & y_{22} & \dots & y_{2j} & \dots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ y_{i1} & y_{i2} & \dots & y_{ij} & \dots & y_{in} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \dots & y_{nj} & \dots & y_{nn} \end{bmatrix} \quad (2)$$

In Eq. (2),  $\tanh(x)$  shows the neuron activation function,  $x_i$  embodies the  $i$ -th neuron membrane voltage,  $Y$  characterizes the synaptic weight matrix, and  $y_{ij}$  denotes the synaptic weight between  $j$ -th and  $i$ -th neurons. Moreover,  $I_i$  characterizes  $i$ -th neuron external stimulate current. In recent times, improved model has been created on the basis of original HNN models, namely HNN with time delay, fractional HNN, discrete HNN, HNN with dissimilar active functions, etc.

The ARO algorithm can be applied to improve the detection rate of the HNN model. The ARO can be stimulated by the survival skills of the rabbit [28]. Rabbits are herbivores which mainly consume leafy weeds and grass. Rabbits wouldn't eat the grass nearby the holes; rather, they find food far away from their nests to avoid predators identifying the nest. These foraging strategies are determined as exploration. Furthermore, to lessen the possibility of being captured by hunters or predators, they are skilled at digging a lot of holes for the nest and randomly choose one as a shelter. This random hiding approach can be assumed as exploitation in ARO. Rabbits must run faster to avoid dangers from the predator owing to their low level in the food chain, resulting in a decline in their energy, so they should shift between random hiding and detour foraging based on their energy status. The mathematical model of ARO is constructed with previous knowledge about the natural behaviors of rabbits, such as exploitation, exploration, and transition from exploration to exploitation.

Consider that every individual in the population has an individual area with burrows and few grass. In foraging activity, the rabbit has a tendency to move towards the faraway area of other rabbits in finding food and overlook what lies nearby, same as an old Chinese proverb says: "A rabbit doesn't eat grass close to their nests". These behaviors are named detour foraging, and they can be mathematically formulated as:

$$X_i(t+1) = X_j(t) + A \times (X_i(t) - X_j(t)) + \text{round}(0.5 \times (0.05 + R_1)) \times n_1, \quad (3)$$

$$i, j = 1, \dots, N \text{ and } i \neq j$$

$$A = L \times c \quad (4)$$

$$L = \left( e - e^{\frac{t-1}{T}} \right) \times \sin(2\pi R_2) \quad (5)$$

$$c(k) = \begin{cases} 1, & \text{if } k == g(l) \\ 0, & \text{otherwise} \end{cases} \quad k = 1, \dots, D \text{ and } l = 1, \dots, [R_3 \times D] \quad (6)$$

$$g = \text{randperm}(D) \quad (7)$$

$$n_1 \sim N(0,1) \quad (8)$$

Where  $X_i(t)$  and  $X_j(t)$  signify the location of  $i$ -th and  $j$ -th rabbits at the  $t$  existing iteration,  $X_i(t + 1)$  indicates the candidate location of  $i$ -th rabbit at  $t + 1$  the next iteration correspondingly.  $T$  denotes the higher iteration counts.  $N$  shows the size of population.  $t$  represents the existing iteration.  $\lceil \cdot \rceil$  refers to the ceiling function.  $D$  symbolizes the dimensional of specific problem.  $randperm(\cdot)$  shows the arbitrary value within 1 and  $D$ .  $R_1$ ,  $R_2$ , and  $R_3$  indicates the random integer within  $[0,1]$ .  $round(\cdot)$  indicates rounding to the nearby integer.  $L$  stands for the length of movement stage while implementing the detour foraging.  $n_1$  follows the uniform distribution.

Here, rabbits tend to conduct continuous detour foraging at the beginning of iteration; then, they often implement random hiding. The idea of rabbit energy  $E$  was introduced to retain a better balance between exploitation and exploration that is gradually reduced over time:

$$E(t) = 4 \left(1 - \frac{t}{T}\right) \ln \frac{1}{R_4} \quad (9)$$

In Eq. (9),  $R_4$  indicates the random integer having range of  $[0, 1]$ . The value of  $E$  energy co-efficient differs from zero to two. If  $E \leq 1$ , it shows that rabbit has lesser energy for physical activities. Hence it is necessary to carry out random hiding to escape from the predators, and the ARO method enters the exploitation stage. If  $E > 1$ , it shows that rabbit has sufficient energy to discover the foraging region of other individuals such that the detour foraging takes place, and this stage can be determined by the exploration. Rabbits are generally met with attack and chase from the hunters. To survive, they should dig several holes nearby their nests for shelter.

In Eq. (9), the variable  $R_4$  represents a randomly generated integer within the range  $[0, 1]$ . The energy coefficient  $E$  assumes values from zero to two. When  $E \leq 1$ , it indicates that the rabbit possesses limited energy for engaging in physical activities. As a result, the rabbit adopts a strategy of random hiding to evade predators, marking the onset of the exploitation stage in the ARO method. Conversely, when  $E > 1$ , the rabbit possesses sufficient energy to explore the foraging regions of other individuals. This condition triggers detour foraging and signifies the exploration stage. In their natural environment, rabbits often encounter threats from predators, leading to pursuits and attacks. To ensure survival, they create several burrows in close proximity to their nests, offering shelter from potential threats.

$$X_i(t + 1) = X_i(t) + A \times (R_5 \times b_{i,r}(t) - X_i(t)) \quad (10)$$

$$b_{i,r}(t) = X_i(t) + H \times g_r(k) \times X_i(t) \quad (11)$$

$$g_r(k) = \begin{cases} 1, & \text{if } k == \lceil R_6 \times D \rceil \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

$$H = \frac{T-t+1}{T} \times n_2 \quad (13)$$

$$n_2 \sim N(0,1) \quad (14)$$

Where the parameter  $A$  is evaluated by Eqs. (4)-(7),  $R_5$  and  $R_6$  shows two random integers within  $[0,1]$ ,  $b_{i,r}(t)$  signify the arbitrarily chosen burrow of  $i$ -th rabbits in  $D$  burrows applied

to hide at  $t$  existing iteration, and  $n_2$  follows the uniform distribution.

Fitness selection is a crucial component of the ARO technique. Encoded outcomes are utilized to assess the quality of solution candidates. In this context, the accuracy value serves as the primary criterion for designing a fitness function (FF).

$$Fitness = \max(P) \quad (15)$$

$$P = \frac{TP}{TP+FP} \quad (16)$$

Where  $TP$  denote the true positive and  $FP$  specifies the false positive value.

### C. Disease Detection using DELM Model

At this stage, the DELM model is used to detect the presence of the disease. ELM is the first presented by Huang et al. that is utilized for SLFNs [29]. An input weighted and hidden layer (HL) biases can be arbitrarily allocated at first, so the trained databases for determining the resultant weighted of SLFNs are integrated. For  $N$  random various instances  $(x_i, t_i)$ ,  $i = 1, 2, \dots, N$ , whereas  $x_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T$ ,  $t_i = [t_{i1}, t_{i2}, \dots, t_{im}]^T$ . Thus, the ELM technique is expressed as:

$$\sum_{j=1}^L \beta_j g_j(x_i) = \sum_{j=1}^L \beta_j g(w_j \cdot x_i + b_j) = 0_i (i = 1, 2, \dots, N), \quad (17)$$

Whereas  $\beta_j = [\beta_{j1}, \beta_{j2}, \dots, \beta_{jm}]^T$  states the  $j^{th}$  hidden node weighted vector, but the weighted vector among the  $j^{th}$  hidden node and the resultant layer is defined as  $w_j = [w_{1j}, w_{2j}, \dots, w_{nj}]^T$ . The threshold of  $j^{th}$  hidden node is expressed as  $b_j$ , and  $0_i = [0_{i1}, 0_{i2}, \dots, 0_{im}]^T$  refers to the  $i^{th}$  resultant vector of ELM.

It is estimated the resultant of DELM when the activation function  $g(x)$  with 0 error that implies as Eq. (18):

$$\sum_{i=1}^N ||0_i - t_i|| = 0. \quad (18)$$

Thus, Eq. (17) is termed as Eq. (19):

$$\sum_{j=1}^L \beta_j g_j(x_i) = \sum_{j=1}^L \beta_j g(w_j \cdot x_i + b_j) = t_i (i = 1, 2, \dots, N). \quad (19)$$

Eventually, Eq. (19) is easily defined as Eq. (20):

$$H\beta = T, \quad (20)$$

whereas,  $H$  defines the HL resultant matrix, and  $H = H(w_1, w_2, w_L, b_1, b_2, b_L, x_1, x_2, x_N)$ . So,  $h_{ij}$ ,  $\beta$ , and  $T$  are demonstrated as:

$$[h_{ij}] = \begin{bmatrix} g(w_L \cdot x_1 + b_L) & \dots & g(w_L \cdot x_1 + b_L) \\ \vdots & \ddots & \dots \\ g(w_L \cdot x_N + b_L) & \dots & g(w_L \cdot x_N + b_L) \end{bmatrix}, \quad (21)$$

$$\beta = \begin{bmatrix} \beta_{11} & \beta_{12} & \dots & \beta_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{L1} & \beta_{L2} & \dots & \beta_{Lm} \end{bmatrix} \quad (22)$$

and



$$T = \begin{bmatrix} t_{11} & t_{12} & \dots & t_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ t_{N1} & t_{N2} & \dots & t_{Nm} \end{bmatrix}. \quad (23)$$

Afterwards, the minimal norm least-squares solution of Eq. (20) as:

$$\hat{\beta} = H^{\dagger}T, \quad (24)$$

Whereas  $H^{\dagger}$  implies the Moore Penrose generalization of the inverse of matrix  $H$ . The resultant of DELM is defined as Eq. (25):

$$f(x) = h(x)\beta = h(x)H^{\dagger}T. \quad (25)$$

From the above mentioned, the procedure of ELM is defined as follows. Initially, DELM is arbitrarily allocated the input weighted and HL biased ( $w_i, b_i$ ). Next, it can compute the HL resultant matrix  $H$  based on Eq. (21). Afterward, by employing Eq. (24), it attains the resultant weighted vector  $\beta$ . Lastly, it classifies the novel database based on the above-trained procedure. Fig. 2 represents the framework of ELM.

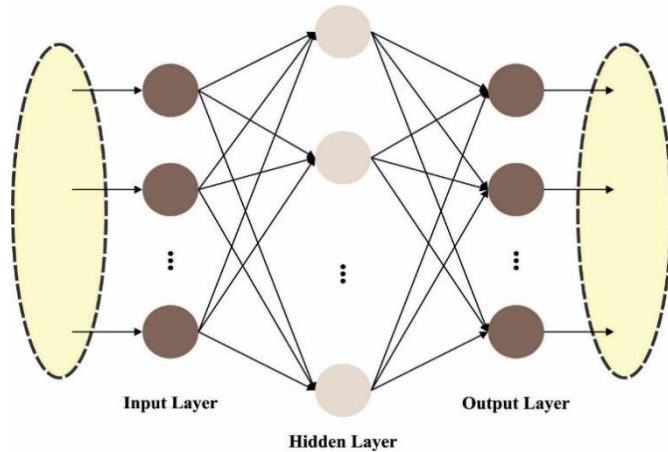


Fig. 2. Architecture of ELM.

#### D. Parameter Tuning using HBO Algorithm

At the final stage, the HBO algorithm was utilized for the optimal parameter tuning of the DELM Model. HBO algorithm was inspired by the social behaviours of human beings according to the hierarchy of organization [30]. This approach stimulates the corporate rank hierarchy (CRH), which implies that the member of the teamwork in the specific organization should be organized in a hierarchical form for completing the specific task. The presented method exploits the CRH model for hierarchically arranging the search candidate according to the fitness of this candidate. For the hierarchy construction, the heap-based data organization can be exploited. Besides the modeling of CRH, the whole concept involves three stages: (i) modeling of the collaborations between their direct manager and the subordinators; (ii) modeling of interactivity amongst the workers; and (iii) lastly, the modeling of self-contribution of the subordinators to accomplish the required task.

#### E. Modelling of the CRH Concept

The presented approach can be conceptualized as a population. In this context, each searching agent within the search space can be likened to a heap node, with the fitness

function (FF) of optimizer problems serving as the master key to access these heap nodes.

In a large organization that operates under a centralized infrastructure, laws and regulations are enforced unilaterally, flowing from senior leadership down to employees. In such a setup, employees are expected to adhere to the instructions of their superiors. With upgrading the place of searching candidate, this stage is mathematically defined:

$$x_i^k(t+1) = B^k + \gamma(2r-1)|B^k - x_i^k(t)| \quad (26)$$

In Eq. (26),  $x$  indicates the position of search agent;  $t$  and  $k$  show the existing iteration and the vector element, correspondingly; and  $B$  shows the parental node. The term  $(2r-1)$  symbolizes the  $k$ -th components of the vector  $\gamma$  and is produced randomly and defined as follows:

$$\lambda^k = 2r - 1 \quad (27)$$

In Eq. (27),  $r$  indicates the arbitrary parameter within  $[0,1]$  in a uniform distribution:

$$\gamma = \left| 2 - \frac{(t \bmod T)}{\frac{T}{4C}} \right| \quad (28)$$

In Eq. (28),  $T$  shows the maximal amount of iterations, and  $C$  indicates an adjustable parameter and relies on the iteration based on Eq. (29):

$$C = \frac{T}{25} \quad (29)$$

Colleagues (Subordinators) in a specific organization cooperate to accomplish official tasks. In the presented method, the nodes at a similar location from the heap are considered colleagues:

$$x_i^k(t+1) = \begin{cases} S_r^k + \gamma\lambda^k|S_r^k - x_i^k(t)|, & f(S_r) < f(x_i(t)) \\ x_i^k + \gamma\lambda^k|S_r^k - x_i^k(t)|, & f(S_r) \geq f(x_i(t)) \end{cases} \quad (30)$$

The self-contribution of every sub-ordinator from the organization was defined as follows:

$$x_i^k(t+1) = x_i^k(t) \quad (31)$$

In this section, the three position updating equation defined in the prior subsection is combined as one formula. A roulette wheel was exploited for making a balance among exploitation as well as exploration stages. The  $P_1, P_2,$  and  $P_3$  probabilities are used for achieving the balance between this phase. An initial probability  $p_1$  can be exploited to update the location of the searching agent from the population and is formulated as follows:

$$P_1 = 1 - \frac{t}{T} \quad (32)$$

The second proportion,  $p_2$  can be evaluated by Eq. (33):

$$P_2 = P_1 + \frac{1-P_1}{2} \quad (33)$$

Lastly, the probability  $p_3$  was evaluated by Eq. (33):

$$P_3 = P_2 + \frac{1-P_1}{2} = 1 \quad (34)$$

$$x_i^k(t+1) = \begin{cases} x_i^k(t), & P < P_1 \\ B^k + \gamma\lambda^k |B^k - x_i^k(t)|, & P_1 < P < P_2 \\ S_r^k + \gamma\lambda^k |S_r^k - x_i^k(t)|, & P_2 < P < P_3 \text{ and } f(S_r) < f(x_i(t)) \\ x_i^k + \gamma\lambda^k |S_r^k - x_i^k(t)|, & P_2 < P < P_3 \text{ and } f(S_r) \geq f(x_i(t)) \end{cases} \quad (35)$$

Where  $p$  shows a random value within  $[0,1]$ .

The HBO technique not only grows a FF to attain higher accuracy of classifier and determines a positive integer to represent the greater efficacy of candidate solutions. The decline of classifier error rate is assumed as FF.

$$\begin{aligned} fitness(x_i) &= ClassifierErrorRate(x_i) \\ &= \frac{\text{no.of misclassified instances}}{\text{Total no.of instances}} * 100 \end{aligned} \quad (36)$$

#### IV. RESULTS

##### A. Results Analysis on Intrusion Detection Dataset

In this section, the intrusion detection results of the ARODL-BSSHS approach were tested on the NSL database [31], including 2100 instances and five classes, as shown in Table I.

TABLE I. DETAILS OF NSL DATASET

Class	No. of Instances
Normal_Class	500
DoS_Class	500
Probe_Class	500
R2-L_Class	500
U2-R_Class	100
<b>Total Number of Instances</b>	<b>2100</b>

Fig. 3 illustrates the classifier outcomes generated by the ARODL-BSSHS technique when applied to the NSL dataset. Figs. 3(a) and 3(b) depict the confusion matrix derived from the ARODL-BSSHS method using a 70:30 split of Training and Testing Data Split (TRP/TSP). The outcomes indicate that the ARODL-BSSHS approach effectively identified and correctly categorized all five classes. Similarly, Fig. 3(c) showcases the Precision-Recall (PR) curve yielded by the ARODL-BSSHS approach. The findings suggest that the ARODL-BSSHS system achieved favorable PR performance across all five classes. Lastly, Fig. 3(d) displays the Receiver Operating Characteristic (ROC) curve resulting from the ARODL-BSSHS technique. This graph highlights that the ARODL-BSSHS approach yielded commendable results, exhibiting superior ROC values for all five classes.

The intrusion detection outcomes of the ARODL-BSSHS technique under 70:30 of TRP/TSS are demonstrated in Table II. The results reported that the ARODL-BSSHS technique recognizes five class labels effectually. For instance, with 70% of TRP, the ARODL-BSSHS technique obtains average  $accu_y$  of 99.73%,  $prec_n$  of 99.19%,  $sens_y$  of 99.18%,  $spec_y$  of 99.83%, and  $F_{score}$  of 99.19%. Additionally, with 30% of TSP, the ARODL-BSSHS method attains average  $accu_y$  of 99.75%,

$prec_n$  of 99.45%,  $sens_y$  of 99.46%,  $spec_y$  of 99.83%, and  $F_{score}$  of 99.45%.

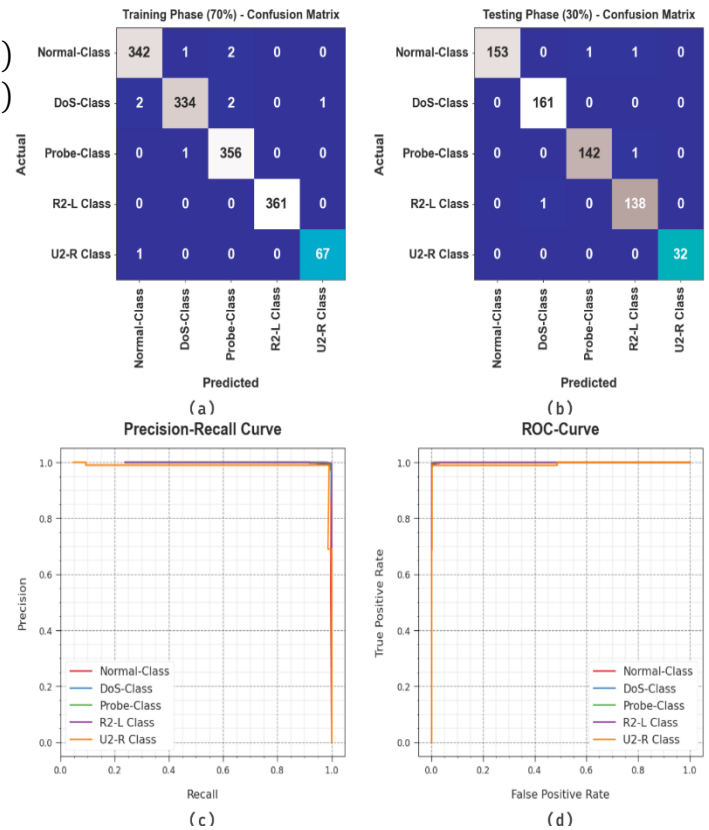


Fig. 3. Classifier outcome on NSL dataset (a-b) Confusion matrices, (c) PR curve, and (d) ROC curve.

TABLE II. INTRUSION DETECTION OUTCOME OF ARODL-BSSHS SYSTEM ON NSL DATASET

Class	$Accu_y$	$Prec_n$	$Sens_y$	$Spec_y$	$F_{score}$
<b>Training Phase (70%)</b>					
Normal-Class	99.59	99.13	99.13	99.73	99.13
DoS-Class	99.52	99.40	98.53	99.82	98.96
Probe-Class	99.66	98.89	99.72	99.64	99.30
R2-L Class	100.00	100.00	100.00	100.00	100.00
U2-R Class	99.86	98.53	98.53	99.93	98.53
<b>Average</b>	<b>99.73</b>	<b>99.19</b>	<b>99.18</b>	<b>99.83</b>	<b>99.19</b>
<b>Testing Phase (30%)</b>					
Normal-Class	99.68	100.00	98.71	100.00	99.35
DoS-Class	99.84	99.38	100.00	99.79	99.69
Probe-Class	99.68	99.30	99.30	99.79	99.30
R2-L Class	99.52	98.57	99.28	99.59	98.92
U2-R Class	100.00	100.00	100.00	100.00	100.00
<b>Average</b>	<b>99.75</b>	<b>99.45</b>	<b>99.46</b>	<b>99.83</b>	<b>99.45</b>

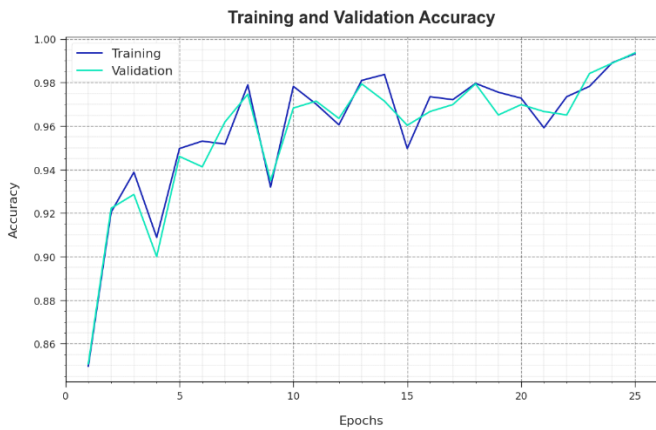


Fig. 4. Accuracy curve of ARODL-BSSHS system on NSL dataset.

Fig. 4 examines the accuracy performance of the ARODL-BSSHS algorithm through the training and validation phases on the NSL dataset. The findings indicate that the ARODL-BSSHS system achieves peak accuracy values as the epochs progress. Notably, the higher validation accuracy in comparison to the training accuracy signifies the proficient learning capability of the ARODL-BSSHS system on the NSL dataset.

The evaluation of loss during both training and validation stages of the ARODL-BSSHS algorithm on the NSL dataset is presented in Fig. 5. The results suggest that the ARODL-BSSHS algorithm maintains similar values of training and validation loss. This observation underscores the effective learning of the ARODL-BSSHS approach on the NSL dataset.



Fig. 5. Loss curve of ARODL-BSSHS system on NSL dataset.

Table III and Fig. 6 reports the comparative intrusion detection results of the ARODL-BSSHS technique. The outcomes implied that the SVM model and LDA model achieves worse outcomes. Although the RF, NB, CART, and HNIDS models offer slightly improved results, the ARODL-BSSHS technique outperforms the other existing models with maximum  $accu_y$  of 99.75%,  $prec_n$  of 99.45%,  $sens_y$  of 99.46%, and  $F_{score}$  of 99.45%.

TABLE III. COMPARISON OF ARODL-BSSHS ALGORITHM WITH DIFFERENT METHODOLOGIES ON THE NSL DATASET

Classification Method	Accuracy	Precision	Sensitivity	F-Score
SVM Model	87.80	97.85	91.41	96.67
LDA Model	91.57	96.78	90.57	96.32
RF Model	97.50	97.85	94.67	97.48
Naïve Bayes	94.66	95.62	93.55	95.08
CART	94.59	96.55	92.78	95.62
LR Model	92.31	91.26	90.11	91.26
HNIDS	98.97	98.85	96.12	99.04
ARODL-BSSHS	99.75	99.45	99.46	99.45

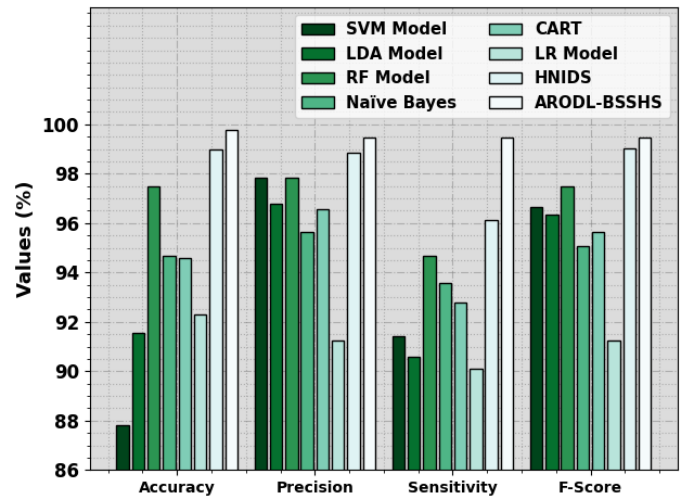


Fig. 6. Comparative outcome of ARODL-BSSHS approach with other methods on NSL dataset.

The computation time (CT) examination of the ARODL-BSSHS technique with recent models on the intrusion detection process is reported in Table IV and Fig. 7. The outcomes reported that the ARODL-BSSHS approach gains least CT of 9.50s. On the other hand, the existing SVM, LDA, RF, NB, CART, LR, and HNIDS models obtain increased CT of 20.54s, 18.89s, 12.37s, 19.77s, 16.09s, 12.97s, and 11.21s respectively.

TABLE IV. COMPARISON OF CT OUTCOME OF ARODL-BSSHS APPROACH WITH OTHERS ON NSL DATASET

Classifier	Computational Time (sec)
SVM	20.54
LDA	18.89
RF	12.37
Naïve Bayes	19.77
CART	16.09
LR	12.97
HNIDS	11.21
ARODL-BSSHS	09.50

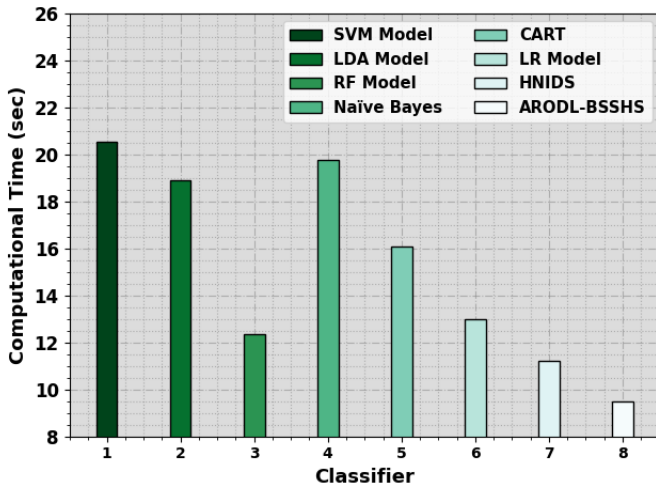


Fig. 7. CT outcome of ARODL-BSSHS approach with other methods on NSL dataset.

B. Results Analysis on Disease Diagnosis Dataset

The Cleveland heart dataset (CHD) [32] contains of 303 samples with 76 features, but only 14 features can be assumed that more appropriate for study experimental purposes. Table V illustrates the details on CHD.

TABLE V. DETAILS ON CHD

Class	No. of Samples
Absence	138
Presence	165
Total Number of Samples	303

Fig. 8 presents the classifier outcomes achieved by the ARODL-BSSHS algorithm when applied to the CHD dataset. Sub-figures 8a and 8b display the confusion matrix generated by the ARODL-BSSHS system using a 70:30 split of Training and Testing Data Split (TRP/TSP). The outcomes indicate that the ARODL-BSSHS system effectively recognized and accurately classified both of the available classes.

Similarly, Fig. 8(c) illustrates the Precision-Recall (PR) analysis performed by the ARODL-BSSHS model. The results reported demonstrate that the ARODL-BSSHS system achieved superior PR performance across the two classes. Lastly, Fig. 8(d) showcases the Receiver Operating Characteristic (ROC) analysis conducted by the ARODL-BSSHS approach. This graph demonstrates that the ARODL-BSSHS algorithm has delivered capable results, achieving maximum ROC values for the two classes.

The classification outcome of the ARODL-BSSHS method under 70:30 of TRP/TSS is established in Table VI. The outcomes stated that the ARODL-BSSHS system recognizes five class labels effectively. For example, with 70% of TRP, the ARODL-BSSHS method attains average  $accu_y$  of 95.07%,  $prec_n$  of 95.32%,  $sens_y$  of 95.07%,  $spec_y$  of 95.07%, and  $F_{score}$  of 95.19%. Furthermore, with 30% of TSP, the ARODL-BSSHS method acquires average  $accu_y$  of 97.83%,  $prec_n$  of

97.87%,  $sens_y$  of 97.83%,  $spec_y$  of 97.83%, and  $F_{score}$  of 97.80%.

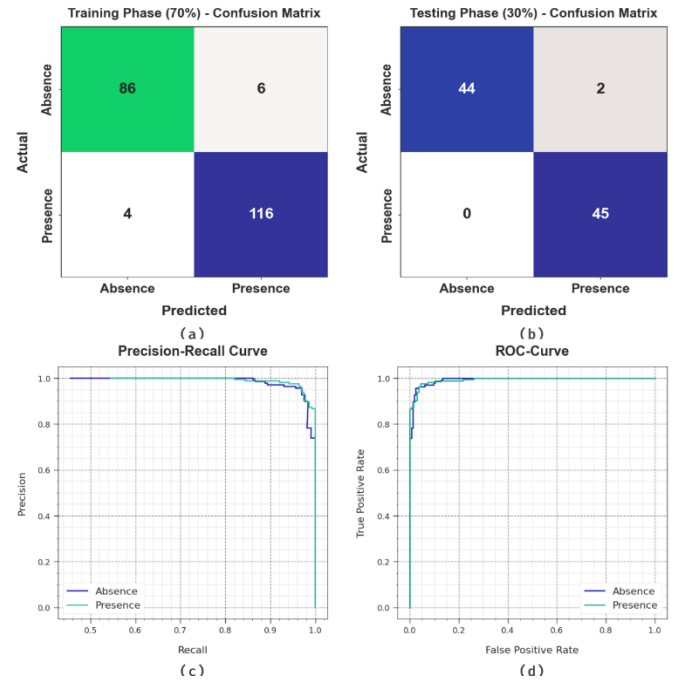


Fig. 8. Classifier outcome on CHD (a-b) Confusion matrices, (c) PR curve, and (d) ROC curve.

TABLE VI. CLASSIFIER OUTCOME OF ARODL-BSSHS SYSTEM ON CHD

Class	$Accu_y$	$Prec_n$	$Sens_y$	$Spec_y$	$F_{Score}$
<b>Training Phase (70%)</b>					
Absence	93.48	95.56	93.48	96.67	94.51
Presence	96.67	95.08	96.67	93.48	95.87
<b>Average</b>	<b>95.07</b>	<b>95.32</b>	<b>95.07</b>	<b>95.07</b>	<b>95.19</b>
<b>Testing Phase (30%)</b>					
Absence	95.65	100.00	95.65	100.00	97.78
Presence	100.00	95.74	100.00	95.65	97.83
<b>Average</b>	<b>97.83</b>	<b>97.87</b>	<b>97.83</b>	<b>97.83</b>	<b>97.80</b>

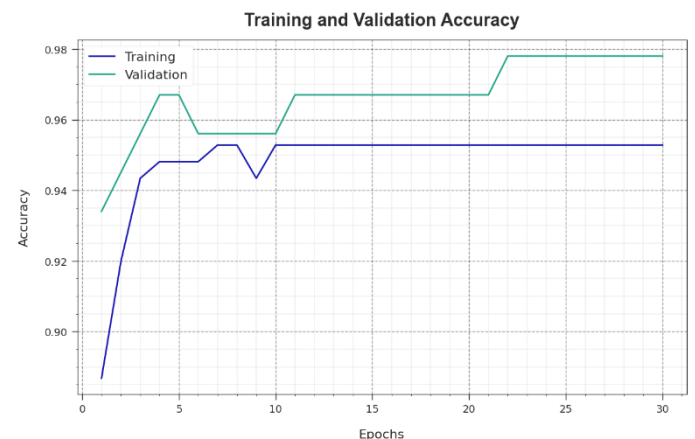


Fig. 9. Accuracy curve of ARODL-BSSHS system on CHD.

Fig. 9 examines the accuracy performance of the ARODL-BSSHS approach within the training and validation phases using the CHD dataset. The results highlight that the ARODL-BSSHS system achieves its highest accuracy values as the epochs progress. Furthermore, the notably superior validation accuracy compared to the training accuracy underscores the efficient learning capacity of the ARODL-BSSHS algorithm on the CHD dataset.

The analysis of loss during both the training and validation stages of the ARODL-BSSHS approach using the CHD dataset is depicted in Fig. 10. The findings suggest that the ARODL-BSSHS algorithm maintains closely aligned values for both training and validation loss. This observation emphasizes the capable learning behavior of the ARODL-BSSHS algorithm on the CHD dataset.

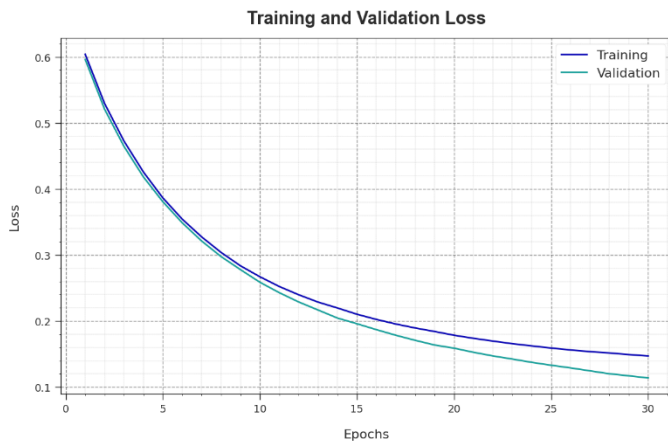


Fig. 10. Loss curve of ARODL-BSSHS system on CHD.

### V. DISCUSSION

In Table VII and Fig. 11, the comparative outcome of the ARODL-BSSHS approach is reported in [33][34]. The results implied that the RF algorithm gains worse performance. But, the NB, LR, SMO, AdaBoostM1 + DS, and Bagging + REPTree approaches offer somewhat higher outcomes; the ARODL-BSSHS system demonstrates the other existing models with maximal  $accu_y$  of 97.83%,  $prec_n$  of 97.87%,  $sens_y$  of 97.83%, and  $F_{score}$  of 97.80%.

TABLE VII. COMPARATIVE OUTCOME OF ARODL-BSSHS APPROACH WITH OTHER METHODS ON CHD

Classifier	Accuracy	Precision	Sensitivity	Specificity	F-Measure
NB Model	84.49	84.50	84.50	87.00	84.50
LR Model	84.49	84.50	84.50	87.00	84.50
SMO Model	86.14	86.20	86.10	90.00	86.10
AdaBoostM1 + DS	83.83	83.90	83.80	88.00	83.80
Bagging + REPTree	83.83	83.90	83.80	88.00	83.80
RF Model	81.19	81.20	81.20	85.00	81.10
ARODL-BSSHS	97.83	97.87	97.83	97.83	97.80

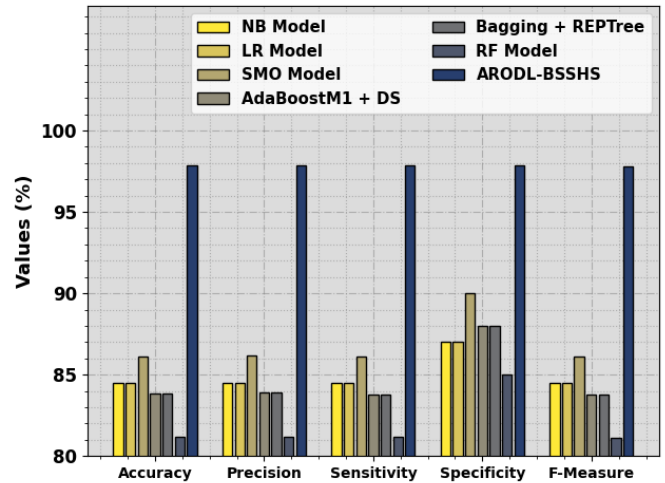


Fig. 11. Comparative outcome of ARODL-BSSHS approach with other methods on CHD.

The CT inspection of the ARODL-BSSHS approach with recent algorithms is reported in Table VII and Fig. 12. The outcomes inferred that the ARODL-BSSHS algorithm reaches a minimal CT of 8.17s. Also, the existing NB, LR, SMO, AdaBoostM1 + DS, Bagging + REPTree, and RF approaches reach maximum CT of 23.20s, 25.10s, 15.90s, 25s, 23.40s, and 20.30s correspondingly. These results analysis assured the better performance of the ARODL-BSSHS technique on the smart healthcare system.

TABLE VIII. CT OUTCOME OF ARODL-BSSHS APPROACH WITH OTHER METHODS ON CHD

Classifier	Computational Time (sec)
NB	23.20
LR	25.10
SMO	15.90
AdaBoostM1 + DS	25.00
Bagging + REPTree	23.40
RF	20.30
ARODL-BSSHS	08.17

The results of the comparative analysis illustrate the superior efficacy of the ARODL-BSSHS approach in securing healthcare systems over the studied alternative models. It was found that ARODL-BSSHS significantly outperforms other classifiers in terms of accuracy, precision, sensitivity, specificity, and F-measure, achieving a maximum accuracy of 97.83% and a minimal Computational Time (CT) of 8.17s. This implies that the ARODL-BSSHS not only is more accurate in predictions and classifications but also is more efficient, making it a preferable choice for real-time applications in smart healthcare systems. This superior performance of ARODL-BSSHS emphasizes the critical role of sophisticated techniques in addressing the complexity and diversity of healthcare requirements and environments. The increased accuracy and reduced computational time are indicative of its capability to deal with the multifaceted and dynamic nature of healthcare data more effectively and

efficiently. The discussed results reinforce the viability and superiority of the ARODL-BSSHS approach in enhancing security and optimizing performance in smart healthcare systems, presenting it as a promising solution for future integrations and developments in healthcare technology.

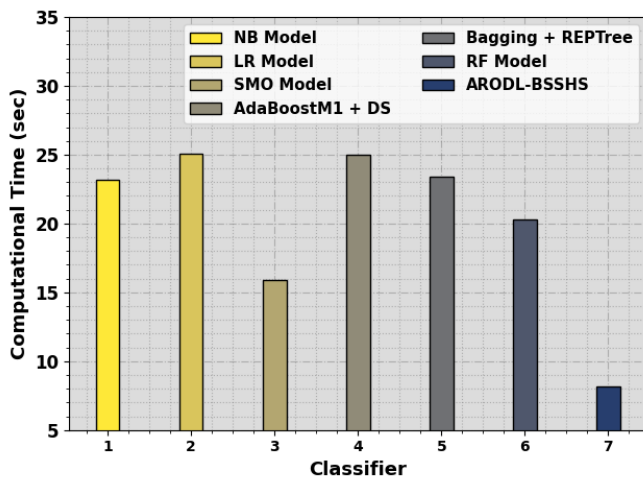


Fig. 12. CT outcome of ARODL-BSSHS approach with other methods on CHD.

## VI. RECOMMENDATIONS & FUTURE WORKS

It is recommended that future research on the ARODL-BSSHS approach should explore its adaptability across diverse sectors like finance, education, and supply chain management. Integration with emerging technologies such as Edge and Quantum Computing and 5G can be crucial to enhance the method's capabilities and to cater to the evolving needs of modern applications. Developing scalable and user-friendly implementations is imperative to ensure broader applicability and user acceptance.

The integration with Edge and Quantum Computing is being explored to optimize computational processes and solve complex problems efficiently [35]. There is also a heightened emphasis on developing robust security and privacy-preserving protocols due to the escalating concerns related to data breaches and cyber-attacks in healthcare systems. The application of Federated Learning and Decentralized AI is gaining traction, addressing the need for decentralized model training and decision-making processes that adhere to data privacy standards. Moreover, the utilization of AI for personalized and predictive healthcare is becoming pivotal, allowing for the development of individualized treatment plans and early detection of diseases.

## VII. CONCLUSION

The ARODL-BSSHS technique has been developed for accomplishing security in the healthcare system in this study. The presented ARODL-BSSHS technique involves the design of secured and smart healthcare system using two major processes, namely intrusion detection and disease diagnosis. To accomplish this, the ARODL-BSSHS technique follows a series of processes: HNN based intrusion detection, ARO based parameter tuning, DELM based disease detection, and HBO based parameter optimization. In addition, the ARODL-

BSSHS technique involves BC technology for secure transmission of healthcare data. A widespread experimental analysis is made on benchmark datasets: heart disease and NSL-KDD dataset to ensure the improved results of the ARODL-BSSHS technique. The experimental values highlighted that the ARODL-BSSHS technique obtains better performance than other approaches. In the upcoming years, the performance of the ARODL-BSSHS algorithm can be improved by multimodal DL techniques.

## REFERENCES

- [1] Ali et al., "Deep learning based homomorphic secure search-able encryption for keyword search in blockchain healthcare system: A novel approach to cryptography," *Sensors*, vol. 22, no. 2, p. 528, 2022.
- [2] H. Bi, J. Liu, and N. Kato, "Deep learning-based privacy preservation and data analytics for IoT enabled healthcare," *IEEE Trans. Ind. Informatics*, vol. 18, no. 7, pp. 4798–4807, 2021.
- [3] P. Sharma, S. Namasudra, R. G. Crespo, J. Parra-Fuente, and M. C. Trivedi, "EHDHE: Enhancing security of healthcare documents in IoT-enabled digital healthcare ecosystems using blockchain," *Inf. Sci. (Ny)*, vol. 629, pp. 703–718, 2023.
- [4] S. Alam et al., "An Overview of Blockchain and IoT Integration for Secure and Reliable Health Records Monitoring," *Sustainability*, vol. 15, no. 7, p. 5660, 2023.
- [5] R. Nishanthini, B. Srimathi, R. S. Kumaran, and I. Yamuna, "Deep Learning on Healthcare Ecosystem using Blockchain Based Security System," in *2021 IEEE Mysore Sub Section International Conference (MysuruCon)*, 2021, pp. 352–357.
- [6] R. Kumar, P. Kumar, R. Tripathi, G. P. Gupta, A. K. M. N. Islam, and M. Shorfuazzaman, "Permissioned blockchain and deep learning for secure and efficient data sharing in industrial healthcare systems," *IEEE Trans. Ind. Informatics*, vol. 18, no. 11, pp. 8065–8073, 2022.
- [7] M. Z. U. Rahman, S. Surekha, K. P. Satamraju, S. S. Mirza, and A. Lay-Ekuakille, "A collateral sensor data sharing framework for decentralized healthcare systems," *IEEE Sens. J.*, vol. 21, no. 24, pp. 27848–27857, 2021.
- [8] M. M. Khubrani and S. Alam, "Blockchain-Based Microgrid for Safe and Reliable Power Generation and Distribution: A Case Study of Saudi Arabia," *Energies*, vol. 16, no. 16, p. 5963, 2023.
- [9] M. M. Khubrani and S. Alam, "A detailed review of blockchain-based applications for protection against pandemic like COVID-19," *TELKOMNIKA (Telecommunication Comput. Electron. Control)*, vol. 19, no. 4, pp. 1185–1196, 2021.
- [10] S. Namasudra, P. Sharma, R. G. Crespo, and V. Shanmuganathan, "Blockchain-based medical certificate generation and verification for IoT-based healthcare systems," *IEEE Consum. Electron. Mag.*, vol. 12, no. 2, pp. 83–93, 2022.
- [11] S. Alam, "The Current State of Blockchain Consensus Mechanism: Issues and Future Works," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 8, 2023, doi: 10.14569/IJACSA.2023.0140810.
- [12] G. A. Rakib et al., "DeepHealth: A secure framework to manage health certificates through medical IoT, blockchain and deep learning," in *2021 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 2021, pp. 1–6.
- [13] H. S. K. Sheth, A. K. Ilavarasi, and A. K. Tyagi, "Deep Learning, blockchain based multi-layered Authentication and Security Architectures," in *2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, 2022, pp. 476–485.
- [14] P. W. Khan, Y.-C. Byun, and N. Park, "IoT-blockchain enabled optimized provenance system for food industry 4.0 using advanced deep learning," *Sensors*, vol. 20, no. 10, p. 2990, 2020.
- [15] G. Zhang, X. Zhang, M. Bilal, W. Dou, X. Xu, and J. J. P. C. Rodrigues, "Identifying fraud in medical insurance based on blockchain and deep learning," *Futur. Gener. Comput. Syst.*, vol. 130, pp. 140–154, 2022.
- [16] A. Lakhani, M. A. Mohammed, J. Nedoma, R. Martinek, P. Tiwari, and N. Kumar, "DRLBTS: deep reinforcement learning-aware blockchain-based healthcare system," *Sci. Rep.*, vol. 13, no. 1, p. 4124, 2023.

- [17] S. K. Singh, Y.-S. Jeong, and J. H. Park, "A deep learning-based IoT-oriented infrastructure for secure smart city," *Sustain. Cities Soc.*, vol. 60, p. 102252, 2020.
- [18] E. A. Mantey, C. Zhou, J. H. Anajemba, I. M. Okpalaoguchi, and O. D.-M. Chiadika, "Blockchain-secured recommender system for special need patients using deep learning," *Front. Public Heal.*, vol. 9, p. 737269, 2021.
- [19] P. Kumar, R. Kumar, G. P. Gupta, R. Tripathi, A. Jolfaei, and A. K. M. N. Islam, "A blockchain-orchestrated deep learning approach for secure data transmission in IoT-enabled healthcare system," *J. Parallel Distrib. Comput.*, vol. 172, pp. 69–83, 2023.
- [20] N. Sammeta and L. Parthiban, "Hyperledger blockchain enabled secure medical record management with deep learning-based diagnosis model," *Complex Intell. Syst.*, vol. 8, no. 1, pp. 625–640, 2022.
- [21] E. M. Abou-Nassar, A. M. Iliyasu, P. M. El-Kafrawy, O.-Y. Song, A. K. Bashir, and A. A. Abd El-Latif, "DITrust chain: towards blockchain-based trust models for sustainable healthcare IoT systems," *IEEE access*, vol. 8, pp. 111223–111238, 2020.
- [22] S. Purbey, B. Khandelwal, and A. K. Choudhary, "Design of a blockchain-based secure and efficient ontology generation model for multiple data genres using augmented stratification in the healthcare industry," *Signal, Image Video Process.*, pp. 1–9, 2023.
- [23] M. A. Almaiah, A. Ali, F. Hajje, M. F. Pasha, and M. A. Alohal, "A lightweight hybrid deep learning privacy preserving model for FC-based industrial internet of medical things," *Sensors*, vol. 22, no. 6, p. 2112, 2022.
- [24] S. P. Dash, "An Introduction to Blockchain Technology: Recent Trends," *Recent Adv. Blockchain Technol. Real-World Appl.*, pp. 1–24, 2023.
- [25] 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.
- [26] S. Alam et al., "Blockchain-based Initiatives: Current state and challenges," *Comput. Networks*, vol. 198, p. 108395, 2021.
- [27] H. Lin et al., "A review of chaotic systems based on memristive Hopfield neural networks," *Mathematics*, vol. 11, no. 6, p. 1369, 2023.
- [28] Y. Wang, Y. Xiao, Y. Guo, and J. Li, "Dynamic chaotic opposition-based learning-driven hybrid Aquila Optimizer and artificial rabbits optimization algorithm: framework and applications," *Processes*, vol. 10, no. 12, p. 2703, 2022.
- [29] S. S. Sammen, M. Ehteram, Z. Sheikh Khozani, and L. M. Sidek, "Binary Coati Optimization Algorithm-Multi-Kernel Least Square Support Vector Machine-Extreme Learning Machine Model (BCOA-MKLSSVM-ELM): A New Hybrid Machine Learning Model for Predicting Reservoir Water Level," *Water*, vol. 15, no. 8, p. 1593, 2023.
- [30] A. S. Menesy, H. M. Sultan, I. O. Habiballah, H. Masrur, K. R. Khan, and M. Khalid, "Optimal Configuration of a Hybrid Photovoltaic/Wind Turbine/Biomass/Hydro-Pumped Storage-Based Energy System Using a Heap-Based Optimization Algorithm," *Energies*, vol. 16, no. 9, p. 3648, 2023.
- [31] 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*, 2009, pp. 1–6.
- [32] J. Andras, S. William, P. Matthias, and D. Robert, "Heart disease. UCI Machine Learning Repository." 1988.
- [33] K. V. V. Reddy, I. Elamvazuthi, A. A. Aziz, S. Paramasivam, H. N. Chua, and S. Pranavanand, "Heart disease risk prediction using machine learning classifiers with attribute evaluators," *Appl. Sci.*, vol. 11, no. 18, p. 8352, 2021.
- [34] A. K. Balyan et al., "A hybrid intrusion detection model using ega-pso and improved random forest method," *Sensors*, vol. 22, no. 16, p. 5986, 2022.
- [35] M. Shuaib et al., "An Optimized, Dynamic, and Efficient Load-Balancing Framework for Resource Management in the Internet of Things (IoT) Environment," *Electronics*, vol. 12, no. 5, p. 1104, 2023

# A QoS-aware Mechanism for Reducing TCP Retransmission Timeouts using Network Tomography

Jingfu LI\*

College of International Education,  
Huanghuai University, Zhumadian 463000,  
Henan Province, China

**Abstract**—A wide range of web-based applications uses the Transmission Control Protocol (TCP) to ensure network resources are shared efficiently and fairly. As wired and wireless networks have become more complex, various end-to-end Congestion Control (CC) schemes have been developed, offering solutions through their proposed TCP variants. Network tomography, a powerful analytical tool, offers a unique perspective by measuring end-to-end performance to estimate internal network parameters, including latency. This estimation capability proves valuable, especially in cases where precise protocol performance evaluation is essential. TCP protocol can be improved significantly by properly estimating RTT time. It has resulted in better network conditions and improved reliability, as well as a higher level of user satisfaction. In this study, we propose a method to infer the link delay using network tomography and then adjust the RTT based on the delay estimation obtained in the previous step. Simulation results performed using the NS2 software show that the proposed method significantly improves the TCP protocol's Round-Trip Time (RTT) estimation by more than 15%. It reduces congestion, improves information transfer efficiency, and ensures the highest level of service in the network.

**Keywords**—Latency; network tomography; end to end; depending on the probe

## I. INTRODUCTION

The landscape of networking technologies has evolved significantly over the years, witnessing advancements that span multiple generations, including the emergence of 5G and the anticipation of 6G. The need for efficient data transfer mechanisms becomes increasingly pronounced as wired and wireless networks expand in complexity and scale. The Transmission Control Protocol (TCP) has long been a cornerstone of network communication, ensuring reliable and adaptive data delivery in various applications [1, 2]. While high-speed networks have reached gigabit speeds, mobile wireless access networks have led to a proliferation of mobile hosts connected to the Internet via slow wireless links [3]. Further, the challenging characteristics of wireless links, such as the high packet loss rates or delays resulting from various factors, such as link-layer retransmissions or handoffs between connection points to the Internet, have posed significant challenges to Internet transport protocols [4, 5]. Today's Internet applications require a reliable mechanism to transfer data due to increasing performance requirements. TCP is extensively used as the transport protocol in many applications due to its ability to adapt to the properties of the network and

its robustness in the face of many types of failures [6, 7]. However, the closed nature of legacy switches, which do not provide accurate visibility of network events, has limited the improvement of the performance of applications that rely on TCP [8].

In this rapidly evolving landscape of complex systems and advanced technologies, research has been directed toward optimizing control strategies and enhancing performance across various domains. The integration of hierarchical optimization and fuzzy logic has yielded promising results in addressing challenges posed by time-varying delays and disturbances in discrete large-scale systems [9, 10]. Machine learning techniques have also emerged as a powerful toolset for analyzing and managing network dynamics. Notably, studies have delved into the analysis of Android ransomware using hybrid approaches [11] and explored machine learning-based network slicing for efficient 5G network management [12]. Additionally, predictive models and probabilistic neural networks have been harnessed to forecast idle slot availability in wireless local area networks [13], while innovative cell designs have been introduced to enhance data collection efficiency in green IoT networks [14]. The integration of machine learning into data conditioning and forecasting methodologies has showcased its potential in optimizing well-pad operations [15, 16]. In the rapidly evolving landscape of complex systems and advanced technologies, the fusion of association rule mining and urban public transportation assumes paramount importance. This synergy facilitates data-driven decision-making, offering insights into commuter behavior, traffic patterns, and service optimization. As urban areas continue to grow, leveraging these tools empowers city planners and transport authorities to navigate the complexities of modern urban mobility, fostering efficiency, sustainability, and improved quality of life for urban dwellers [17, 18]. Moreover, sustainable energy technology has embraced novel concepts, such as power harvesting through ambient vibrations and capacitive transducers [19]. This compilation of research endeavors exemplifies the diverse and impactful directions that contemporary studies are exploring, aiming to unlock new frontiers of knowledge and practical application.

TCP uses the fast retransmit mechanism to trigger retransmissions following the receipt of three consecutive duplicate acknowledgments (ACKs) [20]. As a backoff mechanism, the TCP retransmission timer expires if the TCP sender does not receive ACKs for a certain period of time. Upon expiration of the retransmission timer, the TCP sender



retransmits the first undelivered segment, assuming it is lost in the network [21]. Because retransmission timeouts (RTOs) can indicate that the network is heavily congested, the TCP sender resets its congestion window to one segment and gradually increases it in accordance with the slow start algorithm. Nevertheless, suppose the RTO occurs spuriously, and segments are still outstanding in the network. In that case, a false slow start may damage the potentially congested network by injecting additional segments at an increased rate [22]. One of the most crucial aspects of such a mechanism is how long after sending a package and receiving a receipt, the timeout must be announced. Network tomography offers the opportunity to actively measure link-level characteristics such as delay and loss on end-to-end paths at the link level [23]. Most network tomography techniques developed to date are based on one-way measurements, requiring cooperation from sending and receiving hosts. Consequently, the paths over which these techniques can be applied are severely limited [24]. This paper proposes a method for estimating link delay using network tomography, followed by adjusting the RTT in accordance with the delay value obtained in the previous step. Finally, the results from the study are used to improve the TCP retransmission mechanism to reduce latency and improve performance. The main contributions of this paper are summarized as follows:

- We propose a novel method that leverages network tomography to estimate link delays and adjust RTT values, significantly improving TCP protocol performance.
- Simulation results conducted using NS2 software demonstrate that our approach enhances TCP RTT estimation by over 15%, leading to reduced congestion, improved information transfer efficiency, and heightened service quality within the network.

The remainder of this paper is structured as follows. Section II presents a background of the problem. Section III provides a detailed description of the proposed method, outlining the key steps and innovative aspects of our approach. In Section IV, we present the results of empirical evaluations and comparisons, demonstrating the superiority of our method in terms of accuracy and efficiency. Finally, Section V summarizes our findings, highlights the contributions, and suggests potential avenues for future research.

## II. BACKGROUND

Performance parameters within a network are difficult to measure directly by internal nodes as some may not communicate proactively for security reasons. The network tomography method provides a very convenient means of measuring network parameters since it measures only the end-to-end characteristics of the network. In the end, the network is estimated based on its internal features. Many approaches are employed in network tomographies, such as estimating link level 1 parameters, determining network correlations, and calculating the network traffic matrix between the source and destination networks.

Several methods are presented for estimating link delay via a tomography network. In some cases, the delay distribution is

discrete. These methods assign end-to-end latencies and delay links based on assumed collections, and the probability of mass-delay links is calculated using the EM algorithm [25]. In a statistical model containing latent dependent variables that are imperceptible, the EM algorithm or maximum expected (expectation-maximization) is a repeatable method for finding the maximum value or estimate of the most likely value of an inductive (inferring from effects to causes) parameter set of parameters. The calculation will be repeated for most parameters in the expected results. Calculations and estimates are obtained in stage M to determine the distribution of latent variables used in the next iteration (E) [26]. The method used to estimate the delay of end-to-end latencies supposed links to collections are assigned. The possibility of mass-delay links is calculated using the EM algorithm. There are two limitations in the face of real networks:

- In a real network, because the traffic load on some of the links is heavier and lighter for the other part of the link, packet delay can be quite different. So, if you set that delays end-to-end, links have been assigned to the minor premise. In this case, the desired levels of accuracy are not achieved. Suppose these sets are assumed to be large. In that case, complex computational problems arise, so selecting the appropriate set to assign end-to-end latency and link latency is impossible for these methods.
- The large network size, usually with large amounts of complex data (sets assumed that the end-to-end delays and delays links are assigned to them) to obtain delays of all links face. However, the EM algorithm calculates the probability of each of the probe packets. However, if the input data set is large, these are time-consuming calculations and thus cannot link estimates on time delays for large-scale networks [27].

Unlike the discrete distribution models that estimate the delay, some methods are used to model continuous distribution delay. These methods assume that the distribution of the known delay follows. Examples of distributions include the Poisson distribution and Gaussian distribution compound [26, 28]. Peterson and Davie [29] The authors sought to improve the algorithm to Kaczmarz to take advantage of computing linear tomography systems in the network. The algorithm in linear systems, adaptive control, and tomography systems with wireless sensor networks is used in nature and later as an iterative algorithm in these cases. This algorithm is based on mathematics, potential events, and the full and complex.

In [30], routing and tomography have been amended using a series of evolutionary and biological matrices added and the idea of using algorithms. The framework of evolutionary algorithms solving the robustness and accuracy helps tomography. More studies in the area of network tomography are based on parametric models. The parametric model assumes that measuring data traffic depends on the number of defined parameters. For example, recent studies estimate the internal delay distribution. The probability mass distribution can be modeled as functions. In this context, parameters and the probabilities associated with each function are likely crimes. Some methods of inference delay focused on multicast

routing. The routing of packets sent from the sender to the recipient during an operation moved. During the probe packets, the paths split apart, double, and multiply [31].

### III. PROPOSED METHOD

With the advent of 5G and the impending rise of 6G, the networking landscape has been marked by transformative capabilities, promising higher data rates, lower latency, and greater network efficiency. These advancements are largely attributed to the proliferation of advanced wireless technologies, the expansion of device connectivity, and the integration of cutting-edge communication paradigms. As the industry shifts its focus toward ultra-low latency, these technologies are poised to reshape the dynamics of data transmission and drive the development of novel applications. In this dynamic context, the accurate estimation of Round-Trip Time (RTT) assumes heightened significance. RTT estimation plays a pivotal role in determining optimal retransmission timeouts, facilitating congestion control, and enhancing overall network performance. While conventional methods have sufficed in earlier network paradigms, the evolving landscape introduces new challenges and opportunities that necessitate innovative approaches.

In order to maximize network efficiency, accurate RTT estimation and optimal RTO quantification are essential. The RTO must be longer than the return time or RTT. The RTT is affected by many factors, such as transmission delay, propagation delay, header processing time, ACK production time, etc. Consequently, RTT is not a static value in real-world environments and will change over time. In this case, the change reflects the conditions mentioned in the examples. The RTO should not be set with more than enough, as this will lead to long delays. Additionally, there is an important question to be addressed: what happens if the route changes? What should be done if the network traffic situation and the status of the intermediate nodes change? It is, therefore, necessary to repeat this estimate regularly. The period can be repeated with an appropriate interval of time, such as 20 milliseconds. In order to send probes, a condition may be set, which will result in re-sending the packets if that condition is met. Network conditions will be estimated correctly again. Knowledge of any link to update, to help without sending the actual data and only based on the information provided by the source node is placed tomography, the parameters needed to set up protocols like TCP.

These parameters may also be estimated based on other ideas for calculating them. By estimating the RTT more accurately, the TCP can retransmit at better times and increase the number of these posts. In order to overcome the problems mentioned above, a method based on the unicast delay estimates tomography is recommended for improving TCP. Tomography is used in this way for both end-to-end and link-to-link tomography. Tomography can be used by any network node to calculate the delay on links, whether it is the transmitter or any other node on the network.

Furthermore, it can assist in improving the routing process. To estimate network delay, we first use tomography. To estimate the number of packet delay probe pairs, unreachable packets are sent from the source node to the destination node.

End-to-end packets are bursts that act as end-to-end communications. A delay equation is developed using the relationship between the delay and the delay link that routes packets of information to the probe. Our solution is derived from equations described in the previous chapter to estimate the quantity delay optimally. The route is calculated based on the number of nodes, and each node will have some delay in packets. In general, course delays are little more than the sum of the delays from link to link. As each node experiences a delay in processing, queuing, or propagation, connections are also created. The decision to send real-time packets may result in incorrect calculations and the loss of influence tomography performed to adjust the network RTT in the TCP protocol when the decision is made to send packets in real-time.

In spite of the fact that the processing delay may seem insignificant at first, it does not have any significant impact on large networks with high traffic, especially when there are a number of intermediate nodes and those with high and low processing power. Each node is given a delay. Additionally, it is also possible to have a significant other since it is always in the real world, with a variety of unforeseen and additional delays. Besides the low estimate of the time required for RTT retransmission, a high estimate of the real-time network and reach pass may help reduce the package's delay. When the estimate for the time is slightly larger, it is better to be low since a low estimate, increased network traffic, and poor network conditions can all contribute to delays. By sending and receiving information, each probe packet can help estimate the exact real delay.

For closed questions, the collection, analysis, and length of time from the time of receiving the low are examined, and the amount of delay is determined based on the link. A more appropriate RTT parameter is determined based on the level of probe tomography accuracy of information and links to the detailed estimate of the actual delay. The RTT value is calculated using probes that measure the RTT by tomography and real delay roadshows, achieving a much more accurate delay than TCP. Finally, we will use the results of this calculation to regulate better and improve our TCP retransmission mechanism. By using the proposed approach, better estimates of RTT TCP can be made, and thus, better performance may be achieved.

#### A. Approach Proposed Resolution

As RTT is the basis for accurately estimating probe tomography, multiple packets of burst tomography can be sent to different recipients using unicast protocols. These probes will be used to determine the answer depending on the sender and RTT. The following steps are followed to estimate the actual RTT.

In the first step, tree topology is considered for the network. Leaf nodes of the tree network are selected as the destination. To send pairs of probe packets, multi-packet probe packets are burst using unicast routing. Each packet transmits information about each link, including when it was sent and received by each node, enabling a more accurate estimate of the latency of the links and the route. Data collection and analysis of transmitter probe packets and the amount of time calculated by the probe packet are low. The amount of packet delay is

calculated according to the network. During the extraction and analysis process, we provide accurate, real-time probe data packet processing at each node to estimate the actual RTT at each node. The information obtained from the probe packets is used to calculate the actual delay based on the optimal number of links as follows:

$$\text{Path Delay} = (\text{Tomo Delay}) + ((2n-2) * (\text{Mean Node Delay})) \quad (1)$$

In Eq. (1), Tomo Delay is the optimal amount of delay quantity for links through tomography, n is the number of nodes between source and destination, and mean Node Delay represents the average delay of processing nodes. The probability does not decrease if the sweep is different and has a different number of nodes. If there were more nodes along the path, the equation was recalculated. By replacing parameter 2n with the number of nodes and the number of nodes in the path, the equation is modified as follows:

$$\text{Path Delay} = (\text{Tomo Delay}) + ((n_{\text{went}} + n_{\text{back}}) * (\text{Mean Node Delay})) \quad (2)$$

To improve the mechanism of retransmission of TCP, the result calculation shows that the actual delay replaces the RTT value calculated by the TCP protocol so that TCP will begin with a more accurate estimate of RTT. In practice, more complete equations may be available instead of replacing the equation. There is, however, an additional condition that can be met by a simple equation to estimate the actual delay. The routine can be effective for a network. A better match will always be required in computational experiments.

### B. Evaluation of the precision of the delay

This section provides information regarding the delay time, accuracy, and reliability. In addition to accurately representing scientific work, it indicates what they can contribute to other research endeavors. Estimated delay in TCP is considered an important issue, and its accuracy is crucial. We use the cumulative method to adjust and improve TCP RTT. If the estimates are correct, they will have a significant impact on the

performance of the network. Statistical analysis indicates that the proposed method effectively predicts the delay, and the results indicated they would be considerable. It is, therefore, necessary to compare the delay used to calculate the cumulative delay in the proposed method with the actual delay by the time an estimate is made regarding the links and integrated.

The complexity and variability of the network make it impossible to test with a definitive result. So, the effort involved here has been estimated at ten times. This means that every time, a predetermined scenario is applied to every package in exchange for the time a probe is sent, calculated, printed, and recorded at the outlet. Additionally, a comparison was made between the sent and received packets obtained and recorded, simultaneously with the actual delay time on the printed output. Because the two packages are sent on two separate paths, each package will experience two delays. Therefore, the two delays are calculated every time. The third scenario is somewhat busier in the network regarding the calculation. The traffic in this scenario consists of 21 TCP streams and 161 on-off Poisson streams transmitted over UDP. Statistics are collected from node one. It can be said that the accuracy of this method is relatively high, and errors in the tenth milliseconds. For more detailed calculations, each row of accuracy and the error rate can be calculated and then averaged. The packet error rate refers to the number of packets with errors divided by the number of packets transmitted, calculated by Eq. (3).

$$PER = \frac{\text{number of packet with errors}}{\text{number of packets transmitted}} \quad (3)$$

Table I provides a comparison of the actual delay and delay calculation for two batches of the probe. This data illustrates the accuracy of our proposed method in estimating delays accurately. Similarly, Table II presents the percentage of delay calculation error in the proposed method for two packs of probe data. Analyzing this data allows us to assess the precision of our approach in real-world scenarios.

TABLE I. COMPARISON OF ACTUAL DELAY AND DELAY CALCULATION FOR TWO BATCHES OF THE PROBE

Closed computational delay the first	Real delay packet first	Delay calculation package II	Real delay packet second
42/5 ms	23.5 ms	10.6 ms	18.6 ms
34.6 ms	87/6 ms	32.7 ms	92/6 ms
5.55 ms	90/5 ms	21/6 ms	7.17 ms
89.7 ms	11.8 ms	29/4 ms	80.4 ms
24/8 ms	91/7 ms	81.7 ms	63/7 ms
94/4 ms	18.5 ms	23.7 ms	65/7 ms
34.5 ms	49/5 ms	48/6 ms	11.7 ms
87.5 ms	08/6 ms	76/3 ms	45/3 ms
12.6 ms	67/6 ms	22/8 ms	59/7 ms
34.8 ms	79/7 ms	34/7 ms	8.57 ms

TABLE II. PERCENTAGE OF A DELAY CALCULATION ERROR IN THE PROPOSED METHOD FOR TWO PACKS OF PROBE

Percent packet error First	Packet error percentage of the second
60/3 percent	20.1 percent
70/7 percent	70/5 percent
90/5 percent	30/13 percent
71/2 percent	80.4 percent
60/10 percent	30.2 percent
60/4 percent	40/5 percent
70/2 percent	80/8 percent
40/3 percent	90/8 percent
20.8 percent	30.8 percent
00/7 percent	40.6 percent

### C. The potential of network tomography-based RTT estimation

In this backdrop, our proposed method offers a novel approach to RTT estimation, leveraging the power of network tomography. Traditional RTT estimation methods may confront limitations in accurately gauging RTT in scenarios with multiple physical paths, varying link delays, and changing network dynamics. Here, network tomography emerges as a promising avenue to address these challenges by providing end-to-end performance insights based on the measurement of delay characteristics. Our method, grounded in network tomography, introduces an additional dimension to RTT estimation. While advancements in networking technologies, such as 5G and 6G, and increased compute power contribute to RTT reduction, our method augments these efforts by enhancing accuracy and reliability. By considering link delays across multiple paths and leveraging network tomography's capabilities, our method provides a more comprehensive and nuanced RTT estimation. Fig. 1 shows the flowchart for actual RTT estimation.

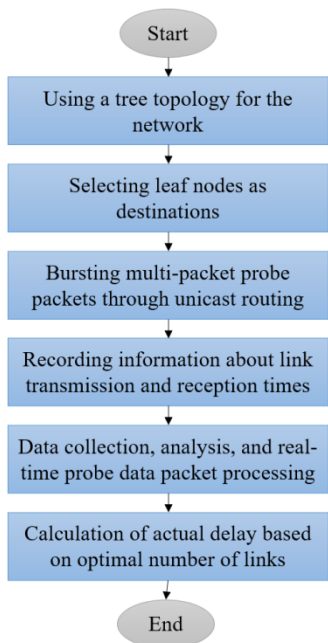


Fig. 1. RTT estimation.

### D. Relevance in evolving networking technologies

The evolving networking technologies, particularly 5G and the anticipated 6G, demand precision in RTT estimation to harness their full potential. While these technologies offer low-latency communication, our method's emphasis on accurate RTT estimation aligns with the quest for precision in data transmission. In environments where real-time applications, IoT devices, and mission-critical communications are paramount, our method's ability to adapt and estimate link delays across diverse paths becomes indispensable. In scenarios with multiple physical paths, MPTCP utilization, and the intricate interplay of network conditions, our method demonstrates its relevance by providing insights into delay characteristics that traditional methods may overlook. This nuanced estimation approach aligns well with the objectives of evolving networking technologies to minimize latency and optimize data transfer efficiency.

## IV. EXPERIMENTAL RESULTS

Simulation may take into account particular circumstances. Simulating and evaluating all possible scenarios is neither feasible nor logical. In each simulation, the actual network conditions should be considered, but the results should be reported as accurately as possible under the constraints. Furthermore, many scenarios are proposed for simulating various aspects of the proposed method. Simulations must also be conducted under conditions similar to those encountered in the review, compared with the method before using it or other suggested methods. In some contexts, a comprehensive network may not be necessary and may be subject to performance conditions. The algorithm is not random in nature, and its behavior is uncertain. Moreover, while improving the network from the perspective of all available parameters, other parameters may be affected. This is because certain reconciliation parameters, namely the balance between them, as well as improving all the parameters simultaneously, cannot be achieved.

The NS2 simulator is used to simulate the proposed method. In order to implement several networks, OTcl is used to create scenarios, and C++ code is used to simulate algorithms and protocols. Communication between scripting languages should be considered in this regard. Although the NS2 software is available on Linux, it can be used on Windows

via a software interface called Cygwin, a Linux virtual machine that can run NS2. The proposed method is simulated and evaluated on a wired network at a fixed time. At this time, seven of eight nodes are linked to a tree with a root node number of zero. Fig. 2 shows this package.

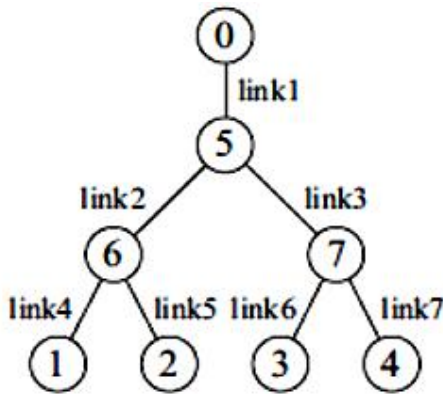


Fig. 2. A network of assessment.

Each link has a queue that is a FIFO algorithm. The simulation was performed with three types of traffic. Traffic is the most crucial first tomography is used, TCP traffic that RTT adjustment is considered. Two types of background traffic are assumed as the frequency of use. CBR traffic, Poisson traffic, and others with entirely different features. Poisson traffic is a traffic accident, while CBR traffic congestion is at the rate specified and is absolutely certain. The CBR traffic intended as a background other than CBR traffic that was initially closed probe is used to estimate RTT TCP traffic. The proposed and other methods and evaluations were conducted in two different scenarios.

#### A. Network Topology

The experimental setup involves a wired network comprising nodes connected in a tree topology, with a root node numbered zero. The nodes are connected through links equipped with a First-In-First-Out (FIFO) queuing algorithm. Two distinct scenarios are considered, each reflecting different aspects of network dynamics and traffic patterns.

Scenario 1 involves a multi-traffic flow setting, where TCP flows are established at rates of 21 units. These TCP flows coexist with background traffic consisting of CBR traffic and Poisson traffic, which is modeled using the UDP. This scenario creates a dynamic and diverse network environment by combining TCP flows with different rates and various types of background traffic, such as CBR and Poisson traffic. The primary objective of this scenario is to thoroughly evaluate and assess the adaptability of the proposed method under conditions of varying network loads and diverse traffic types. Scenario 2 explores the interplay between CBR and TCP dynamics, where TCP flows operating at rates of 21 are integrated with both CBR traffic and background Poisson traffic. The emphasis transitions towards evaluating the effectiveness of the suggested approach when contending with

clearly defined CBR traffic, contributing a heightened level of foreseeability to the network ambiance. This particular setting furnishes valuable observations regarding the method's proficiency in situations characterized by more predictable traffic arrangements. A tree topology governs data routing within the simulated network. The rationale behind this choice lies in its ability to capture a simplified representation of network structures, enabling controlled experimentation while providing a foundation for performance evaluation.

#### B. Results and Analysis

The proposed method's performance is evaluated under a network load with a mix of TCP flows, CBR traffic, and Poisson traffic in the first scenario. The goal is to assess how the algorithm adapts RTT adjustments to optimize network performance under diverse conditions. Fig. 3 and 4 illustrate the comparison of throughput and delay across multiple simulations. The results highlight the method's ability to maintain efficient data transfer despite varying traffic types and network load. The second scenario delves into the method's interaction with competing CBR traffic and effectiveness in a more deterministic setting. Fig. 5 and 6 show the throughput and delay comparisons. The method demonstrates its capability to achieve competitive performance under such conditions, albeit with varying degrees of effectiveness compared to scenario 1. The proposed method's performance is evaluated against different traffic types and rates in both scenarios. The nuanced results reflect the method's adaptability to diverse network conditions and the potential to optimize RTT adjustments effectively. The results underscore the method's potential to provide accurate RTT estimations and optimize data transmission despite varying network dynamics. As networking technologies evolve, our method's capacity to enhance RTT estimation remains relevant and valuable, contributing to optimizing data transfer mechanisms in the face of complex and diverse networking scenarios.

Our proposed method for RTT estimation through network tomography offers versatile, practical applications across diverse contexts. In real-time communication systems, such as telemedicine and remote surgery, where low latency is critical for seamless interactions, our method ensures accurate RTT estimation even in complex, multi-path networks. In IoT deployments, where devices often rely on efficient data exchange for timely decision-making, our approach enhances network reliability by precisely estimating RTT, leading to optimized device communication. Furthermore, in vehicular networks, where vehicles exchange safety-critical information, our method's adaptability to changing network conditions ensures reliable RTT estimation, contributing to safer road environments. In cloud computing environments, where data transfer efficiency impacts application performance, our method aids in precise RTT estimation, leading to enhanced user experiences. These use cases highlight the tangible benefits of our proposed scheme, underscoring its potential to revolutionize diverse industries by improving RTT estimation accuracy, network reliability, and overall performance.

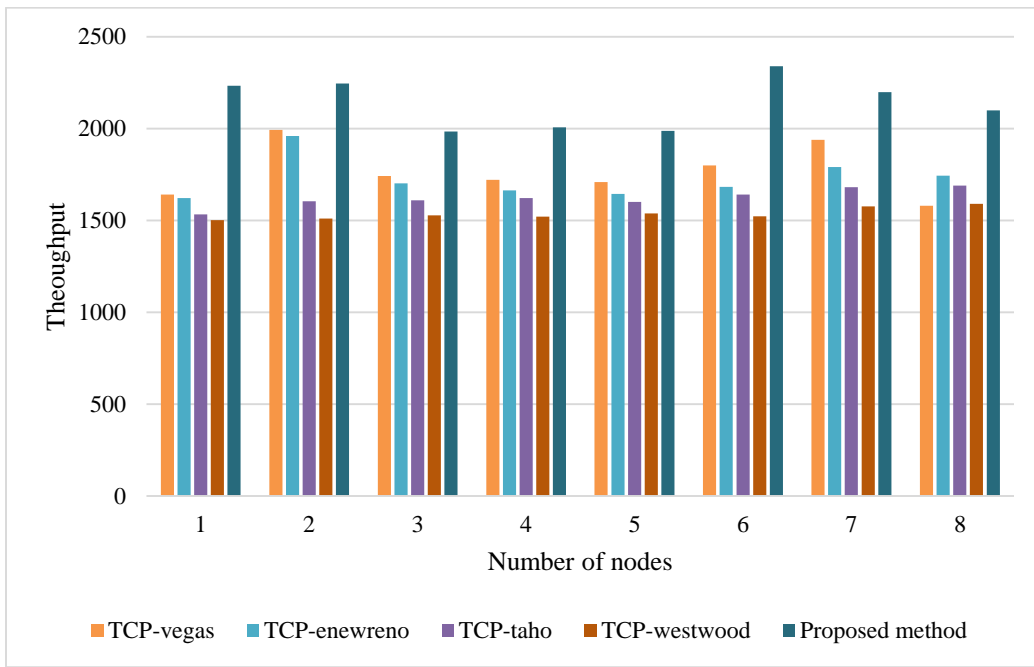


Fig. 3. Comparison of the throughput in the first scenario.

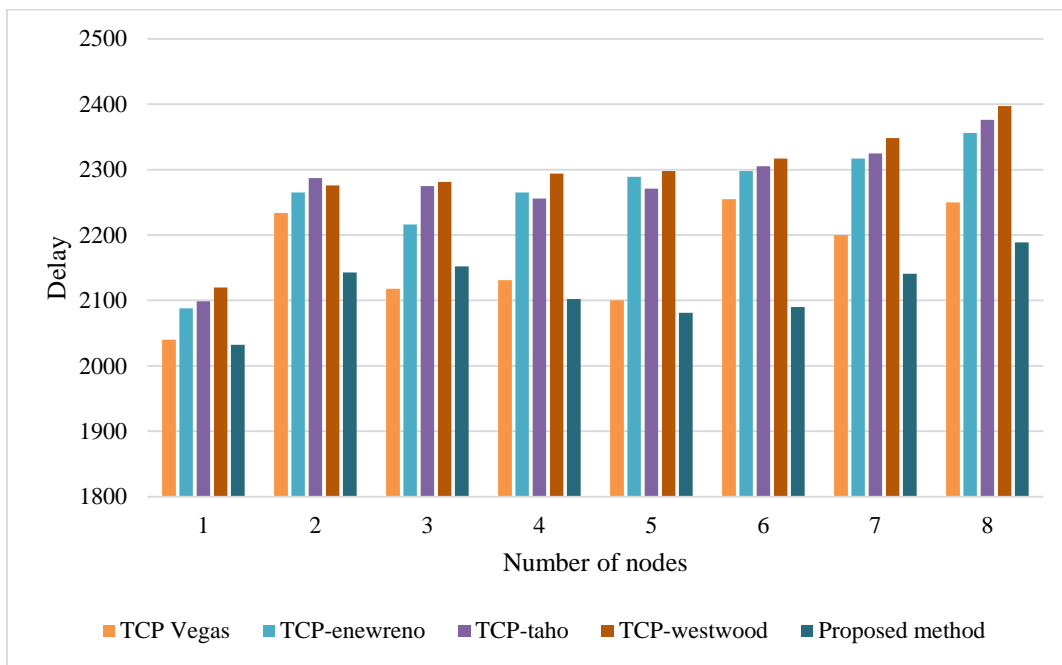


Fig. 4. Comparison of the delay in the first scenario.

While our proposed method significantly enhances TCP's RTT estimation, it is essential to acknowledge that our approach primarily focuses on improving RTT accuracy and reducing congestion in the absence of packet corruption. In cases where packet corruption occurs during transmission, the TCP protocol is equipped with error detection mechanisms, such as checksums, to identify and request retransmission of corrupted packets. Our work primarily complements these

existing error detection and recovery mechanisms by providing more precise RTT estimations, which can lead to more efficient congestion control. However, in cases of severe packet corruption or loss, additional mechanisms at higher protocol layers, such as transport layer error correction codes or application-layer retransmission strategies, may be necessary to ensure data integrity and reliability.

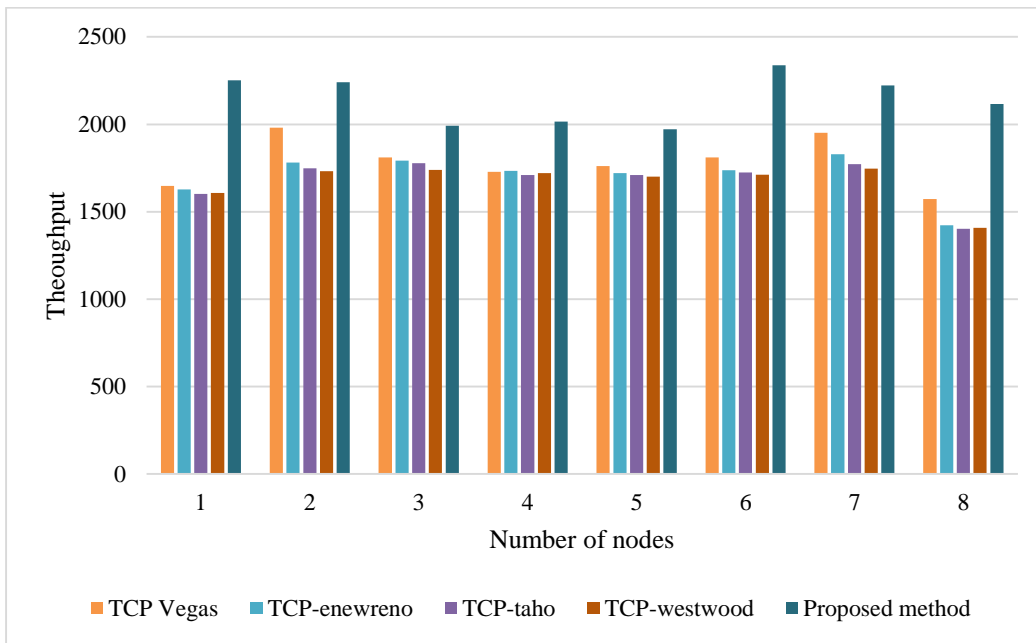


Fig. 5. Comparison of the throughput in the second scenario.

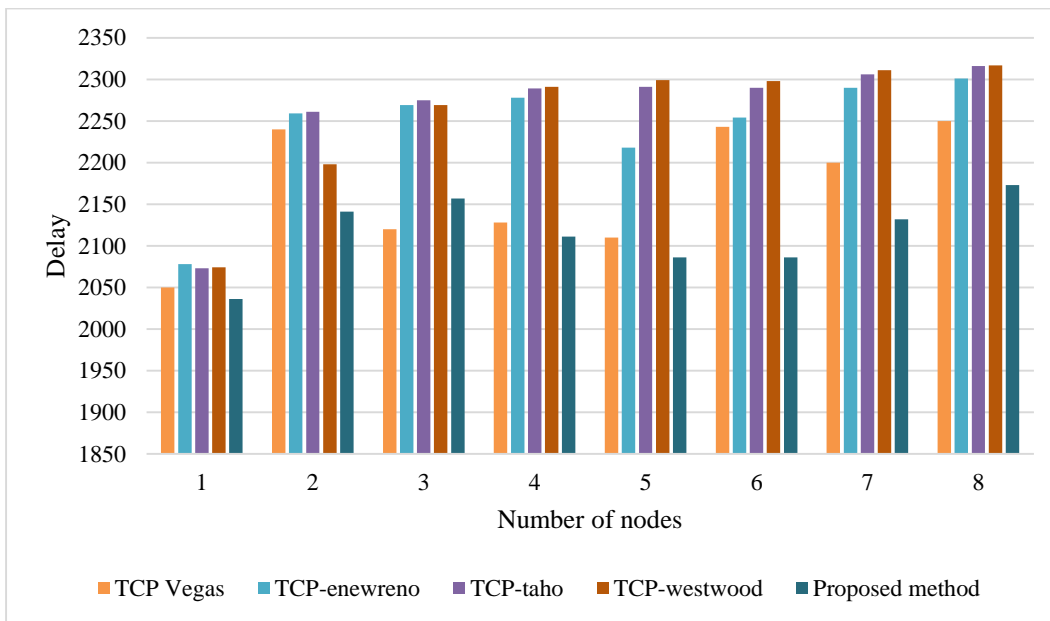


Fig. 6. Comparison of the delay in the second scenario.

## V. CONCLUSION

Network tomography offers an invaluable method for analyzing a network's performance. To determine internal performance parameters, network tomography measures end-to-end performance, as opposed to methods that are based on internal communication. This method has the capability of estimating times in the event that it is necessary to evaluate a protocol's performance in a network by estimating times. The TCP protocol can be significantly improved by properly estimating the RTT time. The proposed method is a relatively flexible method that can be performed using different settings and conditions, and other formulas receive different results. The proposed method is a valuable mechanism for different

computer networks to estimate the rate of actual delay, which deals with network conditions and restrictions to ensure the quality of their service is targeted.

While our method significantly improves RTT estimation, it may not achieve absolute accuracy due to factors such as network variability and unforeseen delays. The sensitivity of our approach to rapid network dynamics necessitates periodic RTT updates, but it may not capture instantaneous changes. Moreover, computational complexities and resource requirements should be considered in resource-constrained environments. Real-world networking scenarios can introduce challenges beyond modeling capabilities, and the scope of our

empirical evaluations may not cover all possible conditions. These limitations provide valuable context for the application and interpretation of our research findings.

#### ACKNOWLEDGMENT

This research is funded by the Henan Province Science and Technology Project (grant No.232102210117,192102210285, 152102210023) and the Research Project of Higher Education Teaching Reform of Huanghuai University (grant No. 2021XJGLX51), and Henan Engineering Research Center of Big Data Analysis and Application for Equipment Manufacturing Internet of Things.

#### REFERENCES

- [1] J. Khan et al., "SMISH: Secure surveillance mechanism on smart healthcare IoT system with probabilistic image encryption," *IEEE Access*, vol. 8, pp. 15747-15767, 2020.
- [2] J. Khan et al., "Efficient secure surveillance on smart healthcare IoT system through cosine-transform encryption," *Journal of Intelligent & Fuzzy Systems*, vol. 40, no. 1, pp. 1417-1442, 2021.
- [3] C. Casetti, "5G Standalone Deployments Are on the Rise [Mobile Radio]," *IEEE Vehicular Technology Magazine*, vol. 18, no. 1, pp. 5-11, 2023.
- [4] A. J. Pinheiro, J. d. M. Bezerra, C. A. Burgardt, and D. R. Campelo, "Identifying IoT devices and events based on packet length from encrypted traffic," *Computer Communications*, vol. 144, pp. 8-17, 2019.
- [5] P. Tomar et al., "Cmt-socks and mptcp multi-path transport protocols: A comprehensive review," *Electronics*, vol. 11, no. 15, p. 2384, 2022.
- [6] G.-M. Sung, C.-T. Lee, Z.-Y. Yan, and C.-P. Yu, "Ethernet Packet to USB Data Transfer Bridge ASIC with Modbus Transmission Control Protocol Based on FPGA Development Kit," *Electronics*, vol. 11, no. 20, p. 3269, 2022.
- [7] L. Zong, H. Wang, and G. Luo, "Transmission Control Over Satellite Network for Marine Environmental Monitoring System," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 10, pp. 19668-19675, 2022.
- [8] H. Bennouri and A. Berqia, "U-NewReno transmission control protocol to improve TCP performance in Underwater Wireless Sensors Networks," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 8, pp. 5746-5758, 2022.
- [9] M. Sarbaz, I. Zamani, M. Manthouri, and A. Ibeas, "Hierarchical optimization-based model predictive control for a class of discrete fuzzy large-scale systems considering time-varying delays and disturbances," *International Journal of Fuzzy Systems*, vol. 24, no. 4, pp. 2107-2130, 2022.
- [10] M. Sarbaz, M. Manthouri, and I. Zamani, "LMI-Based Robust Fuzzy Model Predictive Control of Discrete-Time Fuzzy Takagi-Sugeno Large-Scale Systems Based on Hierarchical Optimization and H<sub>∞</sub> Performance," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 30, no. 04, pp. 649-679, 2022.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] T. Taami, S. Azizi, and R. Yarinezhad, "Unequal sized cells based on cross shapes for data collection in green Internet of Things (IoT) networks," *Wireless Networks*, pp. 1-18, 2023.
- [15] 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.
- [16] R. Soleimani and E. Lobaton, "Enhancing Inference on Physiological and Kinematic Periodic Signals via Phase-Based Interpretability and Multi-Task Learning," *Information*, vol. 13, no. 7, p. 326, 2022.
- [17] M. Shahin et al., "Cluster-based association rule mining for an intersection accident dataset," in *2021 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube)*, 2021: IEEE, pp. 1-6, doi: 10.1109/ICECube53880.2021.9628206.
- [18] S. Saeidi, S. Enjedani, E. Alvandi Behineh, K. Tehranian, and S. Jazayerifar, "Factors Affecting Public Transportation Use during Pandemic: An Integrated Approach of Technology Acceptance Model and Theory of Planned Behavior," *Tehnički glasnik*, vol. 18, pp. 1-12, 09/01 2023, doi: 10.31803/tg-20230601145322.
- [19] U. Jamil, M. Sulaiman, N. Ghafoor, M. Malmir, F. Nawaz, and R. I. Shakoor, "Power Harvesting towards Sustainable Energy Technology through Ambient Vibrations and Capacitive Transducers," in *2023 International Conference on Emerging Power Technologies (ICEPT)*, 2023: IEEE, pp. 1-6.
- [20] G. Amponis, T. Lagkas, K. Tsiknas, P. Radoglou-Grammatikis, and P. Sarigiannidis, "Introducing a New TCP Variant for UAV networks following comparative simulations," *Simulation Modelling Practice and Theory*, vol. 123, p. 102708, 2023.
- [21] C. Lim, "Improving congestion control of TCP for constrained IoT networks," *Sensors*, vol. 20, no. 17, p. 4774, 2020.
- [22] M. Nikzad, K. Jamshidi, A. Bohlooli, and F. M. Faqiry, "An accurate retransmission timeout estimator for content-centric networking based on the Jacobson algorithm," *Digital Communications and Networks*, vol. 8, no. 6, pp. 1085-1093, 2022.
- [23] R. Tagyo, D. Ikegami, and R. Kawahara, "Network tomography using routing probability for undeterministic routing," *IEICE Transactions on Communications*, vol. 104, no. 7, pp. 837-848, 2021.
- [24] A. Ibraheem, Z. Sheng, G. Parisi, and D. Tian, "Neural network based partial tomography for in-vehicle network monitoring," in *2021 IEEE International Conference on Communications Workshops (ICC Workshops)*, 2021: IEEE, pp. 1-6.
- [25] L. Breuer and A. Kume, "An EM Algorithm for Markovian Arrival Processes Observed at Discrete Times," in *MMB/DFT*, 2010: Springer, pp. 242-258.
- [26] G. Antichi, A. Di Pietro, D. Ficara, S. Giordano, G. Procissi, and F. Vitucci, "End-to-end inference of link level queueing delay statistics," in *GLOBECOM 2009-2009 IEEE Global Telecommunications Conference*, 2009: IEEE, pp. 1-6.
- [27] G. Fei, G. Hu, and X. Jiang, "Unicast-based inference of network link delay statistics," in *2011 IEEE 13th International Conference on Communication Technology*, 2011: IEEE, pp. 146-150.
- [28] M.-F. Shih and A. Hero, "Unicast inference of network link delay distributions from edge measurements," in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No. 01CH37221)*, 2001, vol. 6: IEEE, pp. 3421-3424.
- [29] L. L. Peterson and B. S. Davie, *Computer networks: a systems approach*. Elsevier, 2007.
- [30] R. M. Castro, M. J. Coates, and R. D. Nowak, "Likelihood based hierarchical clustering," *IEEE Transactions on signal processing*, vol. 52, no. 8, pp. 2308-2321, 2004.
- [31] J. Lin, J. Zhang, and W. Lin, "The improved method of moment for multicast-based delay distribution inference in network tomography," in *2010 2nd International Conference on Advanced Computer Control*, 2010, vol. 4: IEEE, pp. 486-489.



# Routing Strategies and Protocols for Efficient Data Transmission in the Internet of Vehicles: A Comprehensive Review

Yijun Xu\*

School of Automotive and Rail Transit, Nanjing Institute of Technology  
Nanjing, Jiangsu 211167, China

**Abstract**—The Internet of Vehicles (IoV) integrates wireless communication, vehicular technology, and the Internet to create intelligent transportation systems. Efficient routing of data packets within the IoV is crucial for seamless communication and service enablement. This paper provides a comprehensive review of routing strategies and protocols in the IoV environment, categorizing and evaluating existing approaches. Routing protocols are classified, their adaptability is assessed to network variations, and their performance is compared. Insights are drawn from researchers' experiences. The paper offers a taxonomy of routing protocols, highlights adaptability to network conditions, and presents a comparative analysis. Lessons from researchers shed light on practical implications. The review identifies key routing challenges in IoV and provides a valuable resource for understanding and addressing these challenges in future research.

**Keywords**—Internet of things; internet of vehicles; Vehicular Ad Hoc Networks (VANETs); routing; network adaptability; vehicular technology

## I. INTRODUCTION

The increasing number of users has led to a significant expansion of transportation systems in many countries [1]. However, these systems often suffer from inefficiency and high maintenance costs [2]. The global number of vehicles, including commercial and passenger vehicles, has slightly exceeded one billion, according to recent studies. Projections suggest that it will reach approximately two billion by 2035 [3]. In order to address these challenges, the development of Intelligent Transportation Systems (ITSs) aims to enhance traffic monitoring, road safety, and passenger comfort, ultimately reducing accidents [4]. Vehicular Ad hoc Networks (VANETs) are crucial in implementing intelligent transportation systems. VANETs enable real-time traffic information exchange between two modes of communication: Vehicle-to-Roadside (V2R) and Vehicle-to-Vehicle (V2V) [5]. By facilitating the transmission of warning messages and alerts, VANETs assist drivers in navigating through potential hazards [6]. The main goal of VANETs is to reduce travel time, cost, and pollutant emissions, which in turn enhances traffic safety and efficiency [7]. However, despite the potential benefits, modern vehicular networks face several challenges that must be addressed. Challenges are comprised of unstable internet service, personal devices' limited compatibility, commercialization restrictions, constrained processing capability, network architecture limitations, and no cloud

computing services [8]. Addressing these issues is crucial to harness the full potential of VANETs and to ensure the successful deployment of advanced vehicular networks. By overcoming these challenges, the development of intelligent transportation systems can significantly improve the efficiency, reliability, and overall performance of transportation systems worldwide [9].

Data transmission in the IoV is a critical aspect of modern transportation systems, relying on several innovative technologies to optimize operations and enhance efficiency. Smart grids play a fundamental role by intelligently managing the distribution of energy, enabling Electric Vehicles (EVs) to be a part of IoV seamlessly [10]. Machine learning and deep learning algorithms analyze vast amounts of data generated by vehicles, traffic signals, and urban infrastructure. They derive valuable insights, predicting traffic patterns, suggesting optimal routes, and facilitating efficient energy usage, thereby significantly improving IoV's functionality [11-13]. Artificial Intelligence (AI) acts as the backbone, integrating these technologies and enabling decision-making processes in real time. It enables automated responses to traffic conditions, mitigating congestion and enhancing safety [14-16]. Association rule mining, on the other hand, extracts hidden patterns and correlations from diverse data sources within IoV, revealing valuable information about vehicle behavior, urban mobility, and energy consumption patterns. This knowledge is vital for optimizing routes, managing energy resources, and improving overall transportation efficiency [17]. Urban public transportation is a vital component of IoV, providing sustainable, shared mobility options. Integrating IoV technologies into public transportation enhances services by predicting demand, optimizing schedules, and ensuring a smoother passenger experience. This synergy is paramount in addressing urban traffic challenges, reducing emissions, and transitioning towards smarter, sustainable cities [18].

The IoV network consists of distributed nodes, including vehicles, roadside units, and sensors, which enable local communication. This distributed system facilitates edge computing and the interaction between communication and computation [19]. Artificial Intelligence (AI) is crucial in accessing the IoV network. The IoV network collects information from roadside units and mobile applications, utilizing the bandwidth of the 5G mobile network to enhance internet communication [20]. IoV finds applications in traffic management systems and industrial settings. The future of IoV

research lies in leveraging big data algorithms to process data from IoT devices. It is an evolving field that attracts researchers' attention due to its relevance to human life. Routing protocols specific to the IoV environment are utilized, with SL-ZRP (Stable-Link State Zone Routing Protocol) being a significant communication protocol [21]. SL-ZRP is a function-based protocol that considers factors like speed, destination, and delay to determine optimal routes among vehicles, reducing network representation and overhead. The increasing number of road vehicles poses challenges such as accidents and associated expenses. IoV, originating from VANET, has been the subject of research for several years, addressing these issues. As people's lifestyles change, diverse requirements for vehicular networking have emerged, expanding the scale, structure, and applications of VANET. Large-scale and heterogeneous networks have been introduced, enabling services beyond safety information, including entertainment and environmental protection [22]. This paper proposes a routing protocol taxonomy and explores various IoV applications. This paper makes several significant contributions to the field of IoV:

- **Classification of routing protocols:** The paper provides a comprehensive classification of routing protocols specifically designed for the extreme and complex urban environment of IoV. This classification helps understand the different approaches and strategies employed by these protocols.
- **Adaptability to network density and throughput variation:** The paper recognizes the need for routing algorithms in IoV that can effectively handle low and high network densities while accommodating variations in throughput and delay. This highlights the importance of robust and adaptable routing solutions for the dynamic nature of vehicular networks.
- **Comparison of routing protocols:** The paper offers a comparative analysis of the various protocols in terms of their performance, scalability, reliability, and efficiency. This comparison assists in identifying the strengths and weaknesses of different protocols and aids in selecting the most suitable one for specific IoV scenarios.
- **Lessons learned from researchers:** The paper presents insights and lessons learned from researchers who have explored different challenges related to routing in IoV. This provides a deeper understanding of the practical implications and potential solutions for addressing the unique challenges faced in vehicular networks.

## II. BACKGROUND

### A. Internet of Vehicle

The Internet of Things (IoT) is an evolving technology that links the digital and physical worlds, allowing for communication between objects and humans [23]. This concept has revolutionized our daily lives, making communication more informative, processing more intelligent, and devices smarter [24]. With IoT, the vision of seamless and ubiquitous communication, anytime and anywhere, is

becoming a reality. IoT represents a significant transformation in our lifetime, following the universal accessibility of mobile devices and the world wide web [25]. IoT relies on key technologies such as short-range wireless communications, real-time localization, RFID, and sensor networks. These technologies enable various applications and research areas to flourish, particularly in smart transportation, smart industry, smart homes, and smart healthcare [26]. Integrating smartness in these areas has enhanced efficiency, convenience, and sustainability. Fig. 1 visually illustrates the diverse areas impacted by IoT, highlighting the interconnectedness of smart transportation, smart industry, smart homes, and smart healthcare. The widespread adoption of IoT transforms our environment into a smarter and more interconnected world, revolutionizing how we interact with objects and improving various aspects of our daily lives [27]. The Internet of Vehicles (IoV) is a concept that combines VANET and IoT technologies to establish connections between various devices within vehicles and smart infrastructure on roads [28]. This integration enables seamless communication and data exchange among these devices, leading to a comprehensive IoV-based system [29]. This system includes embedded processors, onboard units, vehicles, roadside units, fog and edge devices, and cloud servers [30]. In an IoV-based system, devices can sense and collect various types of data, such as environmental and traffic-related data. The collected data is then shared among the devices, allowing collaboration and information exchange [31].

Additionally, the data collected from IoV devices can be combined with other data sources, such as social media data, user-generated data, and open-source intelligence, to provide valuable insights for decision-making at different levels [32]. One practical application of IoV-based systems is providing real-time traffic-related information to residents. The system can generate and disseminate up-to-date traffic information by utilizing the data collected from vehicles and roadside units, helping individuals make informed decisions about their routes and travel plans [33].

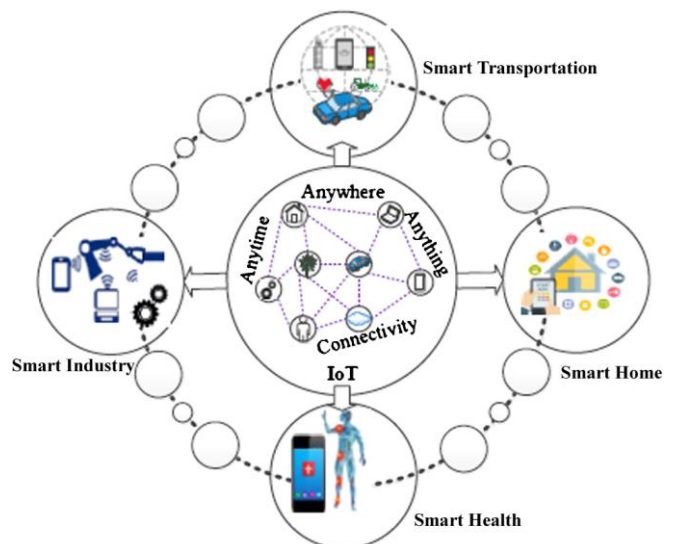


Fig. 1. Diverse areas impacted by IoT.

The IoV is increasingly implemented in urban areas to provide network access to drivers, people, and traffic management personnel. As the transportation system expands, it becomes more challenging and costly to maintain [34]. According to recent reports, the usage of IoV is widespread globally, and it is projected to have over three billion users by 2030. The increased number of vehicles has resulted in traffic congestion and a higher incidence of accidents [35]. To address these issues, IoV is being utilized in urban areas to improve traffic safety. Routing is a crucial aspect of IoV and is vital in daily life. It involves selecting the most optimal path for traffic networks or across multiple networks, considering the dynamic changes in topology [36]. IoV systems detect shortcomings and analyze data to make informed decisions for driving vehicles. Intelligent devices equipped with embedded processors and wireless technologies are utilized in IoV to facilitate vehicle communication [37]. By leveraging various forms of communication, such as device-to-device and machine-to-machine, IoV environments aim to enhance traffic safety in urban areas. Integrating IoV in urban settings aims to reduce traffic accidents by leveraging intelligent technologies and efficient communication. Through real-time data analysis and decision-making processes, IoV systems contribute to improving overall transportation efficiency and enhancing road safety [38].

Incorporating advanced communication and information technology, IoV brings several advantages in resolving traffic and driving challenges, leading to increased passenger safety and a superior driving experience. The communication components of IoV can be categorized into three main types: vehicular mobile Internet, inter-vehicular communication, and intra-vehicular communication [39]. As a heterogeneous vehicular network, IoV involves communication across five different types: Vehicle-to-Infrastructure (V2I), Vehicle-to-Sensors (V2S), Vehicle-to-Personal devices (V2P), Vehicle-to-Vehicle (V2V), and V2R, as illustrated in Fig. 2 [40]. To

facilitate efficient communication in IoV, various wireless technologies are employed. These include vehicular communications such as Dedicated Short-Range Communications (DSRC) and Cellular Automata for Local Mobility (CALM), cellular mobile communication technologies like 4G/LTE, WiMax, and Satellite communication, as well as short-range static communication technologies like Zigbee, Bluetooth, and Wi-Fi [41]. The classification of these wireless communication technologies for IoV applications is depicted in Fig. 3. Fig. 4 provides a general overview of the structure of IoV, illustrating the interconnectedness and communication flow among vehicles, roadside units, sensors, and other components of the IoV ecosystem. This structure forms the foundation for efficiently exchanging information and data within the IoV network.

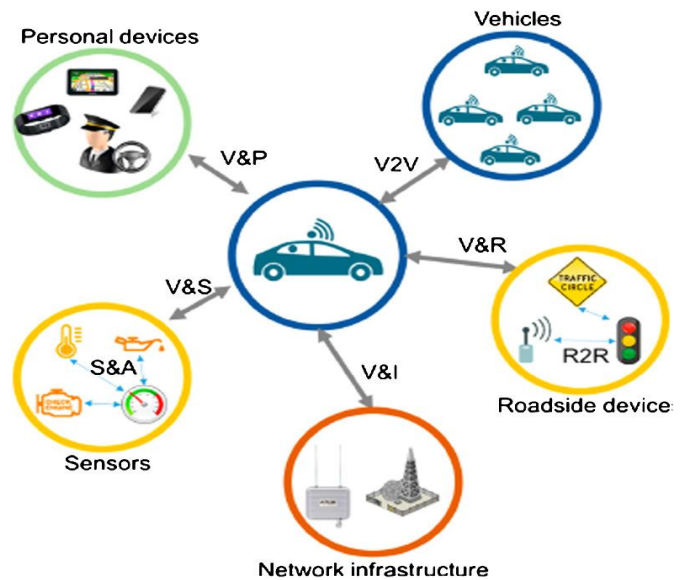


Fig. 2. Communications in IoV.

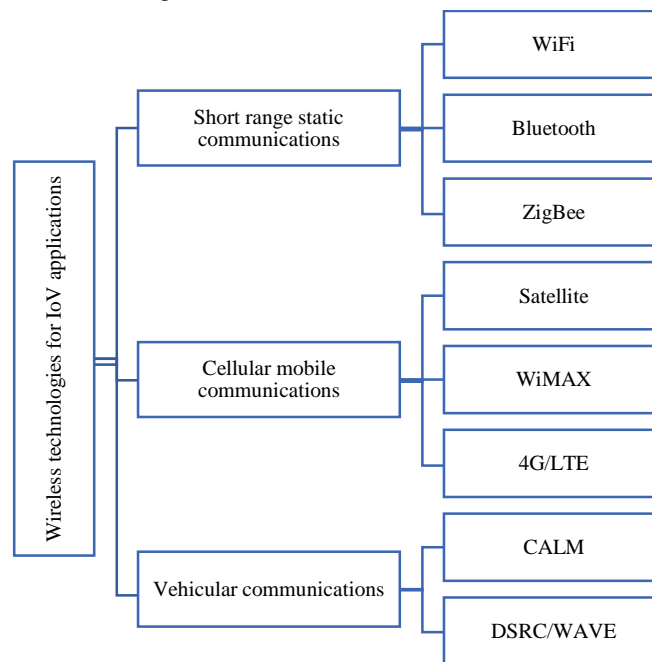


Fig. 3. Wireless communication technologies for IoV applications.

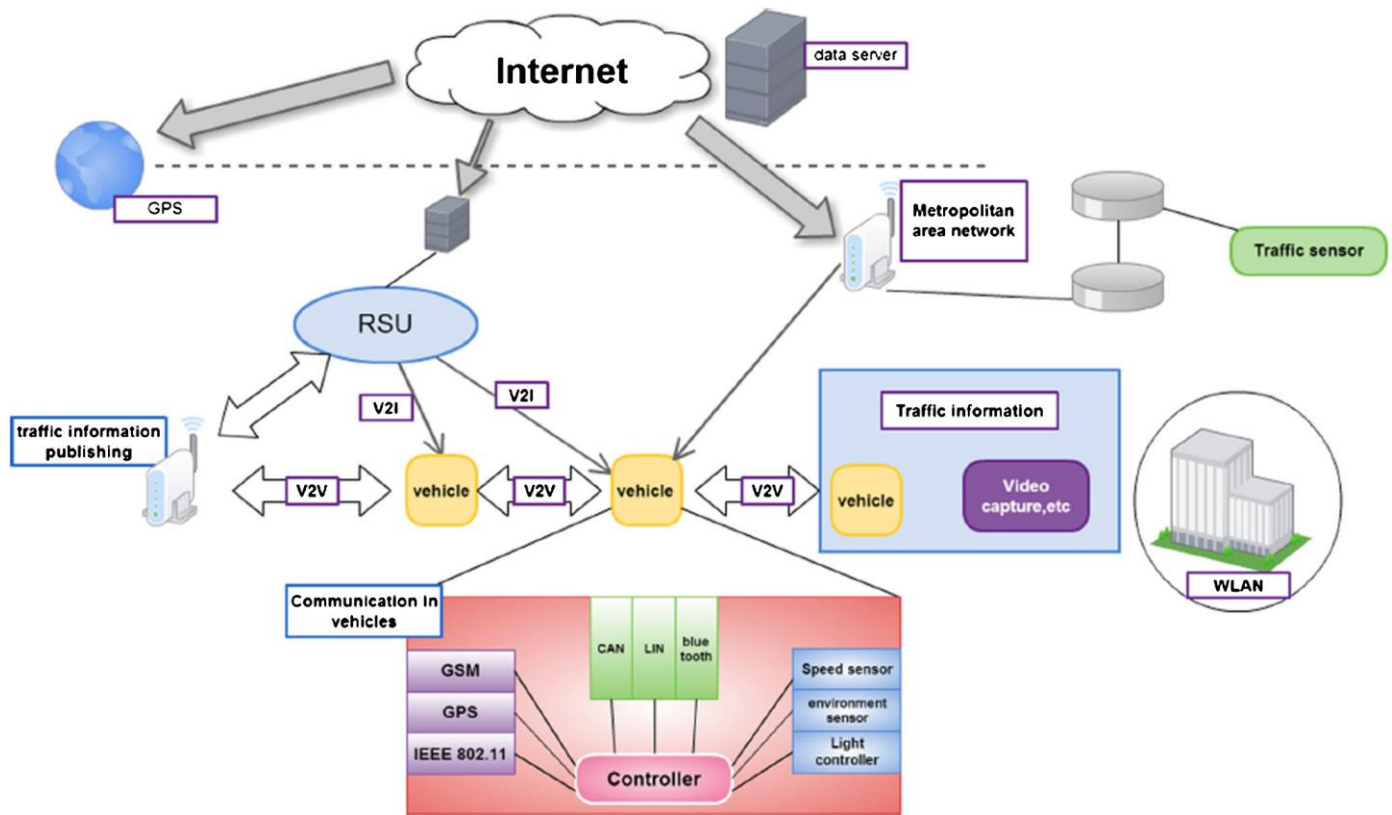


Fig. 4. A general overview of IoV.

The architecture of IoV is structured into four layers: the environment sensing and control layer, application layer, network access and transport layer, and coordinative computing control layer.

- Environment sensing and control layer: This layer plays a crucial role in implementing IoV services by focusing on vehicle control and the traffic environment. It involves sensing and gathering information from the vehicle's perspective and the surrounding environment. Vehicles utilize sensing technology to collect data about the environment, humans, and other vehicles to prevent accidents. Swarm sensing techniques gather dynamic information about the environment and facilitate cooperative decision-making.
- Network access and transport layer: Node management, data processing, remote monitoring, and data analysis are the main tasks in this layer. The IoV network provides every vehicle with diverse network access while taking into account network load constraints. The layer ensures efficient data transmission and handles the transportation of information between vehicles and infrastructure.
- Coordinative computing layer: This layer focuses on coordination within the IoV environment. It supports the interaction of cognitive computing capabilities and swarm intelligent coordinative computing capabilities. The coordinative computing layer facilitates data processing, resource allocation, and decision-making processes within the IoV system.
- Application layer: The application layer offers two types of services: closed and open services. Closed services are specific applications like control platforms and traffic command systems. Open services, provided by numerous internet service providers, include real-time traffic services and must support a suitable business model. Additionally, the application layer enables third-party providers to access open service capabilities, expanding the range of services available within the IoV ecosystem.

#### B. IoV vs. VANET

IoV, an advanced concept that combines VANETs and IoT, aims to enhance the capabilities of VANETs and strengthen ITS. While IoV and VANET technologies aim to improve driving experiences and reduce accidents, several parameters differentiate the two networks. These parameters include their goals, communication types, compatibility, range of usage, processing competence, market attention, network specifications, availability of internet facilities, data size, network connectivity, decision-making processes, the utility of applications, and network awareness [42]. VANETs primarily aim to enhance traffic safety and reduce travel time, costs, and pollutant emissions. However, it lacks entertainment features for passengers, leading to commercialization challenges [43].

On the other hand, IoV technology has broader goals, including improving traffic safety and efficiency and offering commercial infotainment services. IoV's entertainment provides passengers access to online video streaming, movies, file downloading, and other services, thus enhancing their

overall experience. VANETs support two types of communication: V2I and V2V communication [44].

In contrast, IoV enables five types of communication: V2V, V2R, V2I, V2S, and V2P. Each communication type relies on different wireless technologies to exchange information. While individuals widely use personal devices like smartphones, laptops, and tablets, they face compatibility issues within VANETs due to incompatible network architectures [45]. As a result, personal devices cannot effectively communicate information with other nodes in VANETs. In contrast, IoV addresses this compatibility issue, enabling personal devices to efficiently disseminate information among other nodes in the event of hazards, fostering an interactive environment.

The range of usage in VANETs is limited to local and discrete applications, such as providing alerts to drivers about road incidents or avoiding collisions. The nodes in VANETs, which are vehicles, are temporary, random, and unstable, leading to lower scalability compared to IoV [46]. In contrast, IoV offers a global scope and sustainable applications/services by incorporating intelligent vehicular networks with computing and communication capabilities [47]. This enables intelligent networking among vehicles on a larger scale. VANETs face resource constraints regarding computation and processing capacity, as they primarily handle local information collected by sensors in the surrounding environment [48].

In contrast, IoV can handle global information, including big data [49]. Processing and analyzing data in real-time without any delays is crucial. Intelligent computing platforms like cloud computing, fog computing, and edge computing are utilized in IoV for efficient big data analytics and faster processing. VANETs have not achieved the desired commercialization over the years for various reasons, including unreliable internet connectivity, incompatibility with personal devices, and limitations in local processing capabilities. As a result, VANETs have not received significant market attention, and their usage has stagnated. On the other hand, IoV has experienced substantial research advancements and commercial interest. It benefits from reliable internet connectivity, compatibility with personal devices, and the rapid evolution of communication and computation technologies.

VANETs have a singleton network architecture, which limits their usage by not collaborating with other existing networks [50]. This lack of collaboration restricts their connectivity and functionality. In contrast, IoV utilizes a heterogeneous vehicular network framework that enables collaborative networking [51]. IoV incorporates five different types of communications, including WAVE, Wi-Fi, 4G/LTE, and satellite networks, which enhance the flexibility and connectivity of the architecture. Internet connectivity is a fundamental requirement in modern production environments. VANETs face challenges in extending internet connectivity, as roadside infrastructure may be scarce or not fully networked in certain areas.

On the other hand, IoV enables vehicles to connect to the Internet at any time, providing Internet services to all nodes. Faster and reliable internet services in IoV facilitate the implementation of an IoV environment with low latency, high reliability, and increased bandwidth. In terms of data, VANETs

rely on limited local information for decision-making and lack collaboration with global data sources. In contrast, IoV is built on big data principles, as it generates a vast amount of real-time data regarding vehicle information. Additionally, the collaboration among various heterogeneous networks in IoV contributes to accumulating diverse data sources.

Vehicles in VANETs experience frequent disconnections from the ad-hoc network, resulting in a loss of network services. This is primarily due to the non-collaboration with other reachable networks and the pure ad-hoc network architecture. In contrast, vehicles in IoV remain connected to the best available network at all times, enabling efficient communication. IoV can easily collaborate with other reachable networks in case of any issues with the current network. In VANETs, the architecture imposes limitations on storage and computing, making it challenging to make intelligent decisions based on big data mining computations.

On the other hand, IoV architectures leverage Artificial Intelligence-based big data and data mining computations for decision-making. Due to the network disconnection issue in VANETs, the availability of ITS (Intelligent Transportation Systems) applications cannot be guaranteed. In IoV, ITS services are reliable and efficient due to using a client-server architecture with internet connectivity. In VANETs, network services and applications, such as safety messages, require exchanging event location and vehicle information. However, network awareness is limited to neighborhood awareness, as obstacles hinder the proper exchange of information. Additionally, vehicle processing and storage constraints contribute to reduced network awareness. In IoV, incorporating big data technologies like cloud computing and fog computing enhances the network's performance, allowing global network awareness.

### III. CLASSIFICATION OF IOV ROUTING PROTOCOLS

The routing protocols in IoV pose a significant challenge. While many routing protocols, such as DSDV, DSR, and AODV, are adapted from MANET, this article discusses specific geographical routing protocols like GPSR and GPCR. These routing protocols are classified based on their transmission strategies. Location-based routing protocols in IoV can be categorized into four types: hierarchical, geographic, broadcast, and geocast. Among these, geographical protocols are divided into unicast, broadcast, and geocast. The transmission strategy adopted classifies the routing protocols into unicast, geocast, and broadcast categories.

#### A. Transmission Strategy

The transmission strategy in routing protocols for IoV can be classified into three types: unicast routing protocol, geocast routing protocol, and broadcast routing protocol. Unicast routing protocol aims to transmit data from a single source to a single destination using a multi-hop technique through greedy forwarding. Intermediate vehicles can relay the data along a specific routing path from the source to the destination. The routing algorithm determines the forwarding decisions based on specific routing protocol characteristics. Unicast routing protocols can be further categorized based on the information they use into four types: topology-based, position-based, map-

based, and path-based routing protocols. Fig. 5 illustrates the routing strategy and different routing protocols in IoV. The primary objective of geocast routing protocol is to transmit data from a single source node to all destination nodes within a specific geographical region called the Zone of Relevance. It employs a multicast service known as location-based multicast routing. Geocast routing is particularly useful for many VANET applications. Vehicles in the network receive and drop packets based on their current location. Traffic lights can help

implement geocast routing. The broadcast routing protocol is commonly used for sharing information, such as traffic updates, weather emergencies, road conditions, advertisements, and announcements among vehicles. It is also used with unicast routing protocols to discover routes to destinations. Dissemination protocol for heterogeneous vehicular cooperative networks (DHVN) is an example of a broadcast routing protocol. Broadcast routing protocols rely on road topology and network connectivity to function effectively.

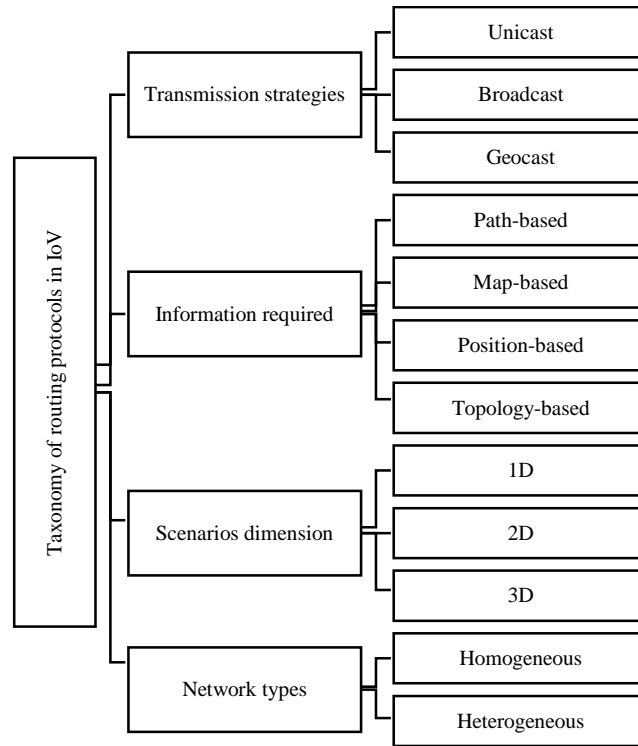


Fig. 5. Taxonomy of routing protocols in IoV.

### B. Information Required

Based on the information required, routing protocols in IoV can be classified into four types:

**Topology-based Routing:** Topology-based routing protocols utilize information about the network's underlying structure, such as the connectivity and topology of the vehicles or infrastructure nodes. These protocols make routing decisions based on the network's topology, including the links between nodes and their quality metrics. Examples of topology-based routing protocols include DSDV (Destination-Sequenced Distance Vector) and OLSR (Optimized Link State Routing). Position-based routing protocols rely on the location information of the vehicles or nodes to make routing decisions. Each vehicle determines its position using GPS or other localization techniques and includes this information in the routing process. Protocols like GPSR (Greedy Perimeter Stateless Routing) and GPCR (Geographic Position-based Routing) fall into the category of position-based routing.

Map-based routing protocols utilize detailed road network maps to make routing decisions. These protocols consider the geographical layout and attributes of the road network, such as road segments, intersections, and traffic conditions. Map data

is used to determine the optimal path for data transmission. Map-based routing protocols are commonly used in navigation and route planning applications. Path-based routing protocols focus on identifying specific paths or routes for data transmission. These protocols use predefined paths or routes, which can be determined based on factors like road conditions, traffic patterns, or specific requirements of the application. Path-based routing allows for more controlled and predetermined data routing. Examples of path-based routing protocols include AODV (Ad hoc On-Demand Distance Vector) and DSR (Dynamic Source Routing).

### C. Scenarios Routing

In addition to the mentioned protocols, there are several other routing protocols in the context of IoV, namely, vehicular routing protocol, Delay Tolerant Routing Protocol (DTN), and Disrupted Adaptive Routing (DAR). Table I shows the classification of IOV routing protocols. Vehicular routing protocol focuses on exchanging road information among vehicles to enable efficient routing. It utilizes techniques like ant colony optimization, where vehicles act as ants to find the optimal path based on local information and pheromone trails left by other vehicles. DTN is designed to handle intermittent or disrupted connectivity in the network. It employs a carry-

forward mechanism, where intermediate nodes store and forward data until a suitable connection becomes available for transmission. DAR is a routing protocol that aims to reduce network congestion and improve overall performance compared to traditional routing protocols. It achieves this by dynamically adapting the routing paths based on network conditions, thereby reducing transmission delays and improving packet delivery.

In IoV networks, routing scenarios are divided into three types: 1D, 2D, and 3D. 1D scenarios involve vehicles moving in a linear direction, such as on a highway or a single-lane road. Routing protocols are designed to facilitate efficient data transmission along this one-dimensional path. In 2D scenarios, vehicles can move in a two-dimensional space, such as urban or suburban areas with multiple lanes and intersections. Routing protocols consider the spatial relationships and connectivity between vehicles in these environments. 3D scenarios involve routing in complex environments where vehicles can move in three dimensions, such as in aerial or underwater vehicular networks. Routing protocols in these scenarios need to account for the specific challenges and characteristics of the respective environments.

#### D. Network Types

Routing is feasible in both homogeneous and heterogeneous networks within the IoV framework. In a

homogeneous network, vehicles and network elements have comparable characteristics and capabilities. The routing protocols for homogeneous networks assume that vehicles have similar communication ranges, transmission capabilities, and network behaviors. Examples of homogeneous networks in IoV include all vehicles equipped with the same communication technology (e.g., all vehicles use Wi-Fi or DSRC for communication). In a heterogeneous network, vehicles and network elements may have different characteristics, capabilities, and communication technologies. Heterogeneous networks in IoV may involve vehicles with different communication ranges, transmission powers, and technologies (e.g., a mix of vehicles using Wi-Fi, cellular networks, or satellite communication). Routing protocols for heterogeneous networks must consider these differences and ensure effective communication and data exchange among vehicles with diverse capabilities.

Routing protocols in homogeneous and heterogeneous networks aim to find the most efficient paths for data transmission, considering factors like network congestion, connectivity, data reliability, and quality of service requirements. The specific design and implementation of routing protocols may vary depending on the characteristics and objectives of the network. Still, the overall goal remains the same: to establish reliable and optimal routes for data transmission in IoV networks.

TABLE I. CLASSIFICATION OF IOV ROUTING PROTOCOLS

Routing protocol type	Information required	Scenarios routing	Network types	Strengths	Examples
Unicast routing	Topology-based	1D and 2D	Homogeneous	Reliable point-to-point communication	DSDV and AODV
	Position-based	1D and 2D	Homogeneous	Efficient use of position information	DSDV and AODV
	Map-based	1D and 2D	Homogeneous	Utilizes detailed road network maps	DSDV and AODV
	Path-based	1D and 2D	Homogeneous	Offers predetermined data routing	DSDV and AODV
Geocast routing	Position-based	2D	Heterogeneous	Efficient data transmission within a specific geographical region	GPSR and GPCR
Broadcast routing	Topology-based	1D and 2D	Homogeneous	Effective for sharing information among vehicles	DHVN
	Map-based	1D and 2D	Homogeneous	Utilizes road network attributes	DHVN
	Path-based	1D and 2D	Homogeneous	Provides controlled data routing	DHVN
Vehicular Routing	Various	Various	Various	Utilizes intelligent algorithms like ant colony optimization	Ant Colony, DTN, and DAR

#### IV. DISCUSSION

In this section, we provide a detailed discussion of the findings and insights gained from the classification of IoV routing protocols. Our comprehensive classification of IoV routing protocols offers a structured approach to categorizing and understanding the diverse strategies employed in vehicular networks. By grouping protocols into three main categories based on their transmission strategies (unicast, geocast, and broadcast), we facilitate a clearer view of their roles and functionalities. This classification provides researchers and practitioners with a valuable roadmap for selecting the most suitable routing protocols for specific IoV scenarios. Unicast routing protocols offer reliable point-to-point communication within IoV networks. Their strengths lie in efficient position-based routing, the utilization of detailed road network maps,

and the ability to provide predetermined data routing. These characteristics make them suitable for various IoV scenarios. Through multi-hop forwarding, intermediate vehicles can relay data, facilitating communication even when the destination is not in the direct transmission range of the source. Unicast protocols are adaptable and can be further categorized based on the information they use, such as topology-based, position-based, map-based, and path-based routing protocols. This flexibility allows for protocol selection that best suits the specific requirements of IoV scenarios. However, these protocols have their limitations. They may face challenges related to network congestion in scenarios with a high density of vehicles. Moreover, data reliability and latency can become concerns in scenarios with dynamic network conditions, necessitating the development of more robust routing algorithms.

Geocast routing protocols are specifically designed to transmit data from a single source to all destination nodes within a predefined geographical region known as the Zone of Relevance. Their strength lies in efficiently disseminating information to a targeted area, making them particularly useful for many vehicular applications. Geocast protocols leverage location-based multicast routing, where vehicles within the designated zone receive and process packets based on their current location. This approach ensures that only vehicles within the relevant geographical area receive the data, reducing unnecessary network traffic. However, geocast protocols have limitations concerning the definition and management of these geographical zones. The accuracy of defining such regions and handling scenarios with overlapping or rapidly changing zones can pose challenges.

Broadcast routing protocols play a crucial role in sharing real-time information among vehicles. Their strength lies in their ability to quickly disseminate critical updates, such as traffic conditions, weather emergencies, or road incidents, to a wide audience of vehicles. Broadcast protocols are often used in conjunction with unicast routing protocols to discover routes to destinations. However, they also have limitations. Broadcasting can lead to network congestion, especially in densely populated areas, where a large number of vehicles simultaneously receive and process broadcast messages. To mitigate this, efficient mechanisms for broadcast suppression and congestion control are necessary. Moreover, ensuring data reliability and minimizing redundant data reception are ongoing challenges in broadcast routing, as data packets may be received by vehicles multiple times.

Configuring routing protocols in the context of the IoV involves a range of parameters that influence how data is routed and communicated within the network. The choice of the routing algorithm is fundamental. IoV can employ various routing protocols, including proactive (table-driven) like OLSR or reactive (on-demand) like AODV. The selection depends on factors like network size, mobility, and application requirements. Parameters related to the network's physical layout, including the number of vehicles, their initial positions, and the road infrastructure, play a significant role. Realistic network topologies are essential for accurate simulations. The communication range of IoV devices, often determined by the technology used (e.g., DSRC, Wi-Fi, cellular), is crucial. It affects how far vehicles can communicate with each other and with roadside infrastructure. Some routing protocols allow for configuring transmission power levels. Adjusting transmission power affects the range at which a vehicle can communicate, influencing network coverage and energy consumption. Parameters related to packet generation, including packet size and transmission rate, can vary depending on the type of data being exchanged. Larger packets or higher transmission rates may require different routing strategies.

Different mobility models, such as Random Waypoint, Gauss-Markov, or real-world traffic data, can be used to simulate vehicle movements. The choice of mobility model affects how vehicles move and interact within the network. In wireless communication, propagation models define how signals propagate through the environment. These models consider factors like path loss, shadowing, and fading,

impacting signal strength and reliability. Parameters related to traffic patterns, such as the type of data generated (e.g., safety messages, multimedia), traffic density, and source-destination pairs, are essential for evaluating routing performance. If the IoV application demands specific QoS, parameters related to latency, jitter, and reliability thresholds may be configured to ensure that routing decisions meet these requirements. Routing protocols in IoV often include security features. Parameters for encryption, authentication, and key management may need configuration to ensure secure communication. Factors like vehicle speed, acceleration, and braking characteristics can be modeled as parameters. These parameters influence how vehicles move and interact in the network. The length of the simulation or data collection period can impact the stability and convergence of routing protocols. Longer simulations may be needed to observe certain network behaviors.

## V. FUTURE RESEARCH DIRECTIONS

- **Standardization and interoperability:** Standardization efforts are crucial to ensure widespread adoption and interoperability of routing algorithms in the IoV. Future research should focus on developing standardized routing protocols and interfaces that enable seamless communication across different vehicular networks and technologies.
- **Blockchain-enabled routing:** Blockchain technology offers decentralized, transparent, and tamper-resistant data management. Integrating blockchain into routing algorithms can enhance trust, security, and privacy in the IoV. Future research should explore the application of blockchain-enabled routing algorithms that can provide secure and reliable communication among vehicles and infrastructure.
- **Machine learning-based routing:** Machine learning techniques have shown promise in various domains. Applying machine learning algorithms to routing in the IoV can enable proactive decision-making, traffic prediction, and congestion control. Future research should investigate the use of machine learning-based routing algorithms that can adapt and learn from network dynamics to optimize routing decisions.
- **Edge computing and intelligent routing:** With the proliferation of edge computing in the IoV, there is an opportunity to leverage edge resources for intelligent routing. Future research should explore intelligent routing algorithms that can utilize edge computing capabilities, such as real-time data processing, decision-making, and resource optimization, to improve routing efficiency and responsiveness.
- **Quality of Service (QoS):** The IoV requires different communication services with varying QoS requirements. Routing algorithms should be able to provide differentiated services based on application-specific QoS metrics, such as latency, reliability, and throughput. Future research should investigate QoS-aware routing algorithms that can efficiently handle diverse application requirements.



- **Security and privacy:** Security and privacy are critical concerns in vehicular networks. Routing algorithms should incorporate robust security mechanisms to protect against attacks and ensure data confidentiality, integrity, and availability. Future research should focus on developing secure routing protocols and privacy-preserving techniques to mitigate threats and protect user privacy effectively.
- **Resilience to attacks:** Vehicular networks are susceptible to jamming, spoofing, and Sybil attacks. Routing algorithms should be resilient to such attacks and capable of detecting and mitigating them. Future research should explore routing algorithms that can enhance the network's resilience and ensure reliable communication in the presence of malicious entities.
- **Scalability:** As the number of connected vehicles increases, scalability becomes a major challenge. Designing routing algorithms that can efficiently handle large-scale networks is crucial. Future research should focus on developing scalable routing schemes that effectively handle increasing vehicles and data traffic.
- **Dynamic network conditions:** Vehicular networks are characterized by high mobility, intermittent connectivity, and dynamic network topologies. Routing algorithms must adapt to these conditions to ensure reliable and efficient communication. Future research should explore adaptive routing algorithms that dynamically adjust routing paths based on the current network state.
- **Energy efficiency:** Vehicles in the IoV are typically resource-constrained, especially in terms of energy. Routing algorithms should consider energy consumption and aim to minimize energy expenditure. Future research should focus on energy-aware routing algorithms that optimize energy consumption while maintaining reliable communication.
- **Integration with smart city infrastructure:** The IoV is closely linked to the concept of smart cities. Routing algorithms should be designed to integrate with existing smart city infrastructure, such as traffic management systems, intelligent transportation systems, and urban sensing networks. Future research should explore routing algorithms that can seamlessly integrate with smart city infrastructure to enable efficient and sustainable urban mobility.
- **Traffic load balancing:** In congested scenarios, routing algorithms should distribute traffic load evenly across the network to prevent congestion and ensure efficient resource utilization. Future research should focus on load-balancing routing algorithms that intelligently distribute traffic and optimize network performance.
- **Vehicular cloud computing:** Integrating cloud computing with vehicular networks offers opportunities for offloading computation and storage tasks to the cloud. Routing algorithms should consider the availability and utilization of cloud resources to optimize communication and resource allocation. Future research should explore routing algorithms that leverage vehicular cloud computing for enhanced scalability, resource management, and application performance.
- **Cross-layer optimization:** Traditional layered network architectures may not be suitable for the dynamic and resource-constrained IoV environment. To improve routing performance, cross-layer optimization techniques can leverage interactions between different layers (e.g., physical, MAC, and network). Future research should investigate cross-layer routing algorithms that optimize communication efficiency, reliability, and resource utilization.
- **Robustness to mobility:** Vehicles in the IoV are highly mobile, resulting in frequent topology changes and link disruptions. Routing algorithms should be robust to mobility-induced challenges and maintain connectivity even in highly dynamic environments. Future research should explore mobility-aware routing algorithms that can adapt to vehicle movements and ensure seamless communication.
- **Cooperative communication and collaboration:** Cooperative communication among vehicles and infrastructure can improve routing efficiency, reliability, and safety in the IoV. Research should focus on cooperative routing algorithms that enable vehicles to collaborate, exchange information, and assist each other in routing decisions, leading to enhanced network performance.
- **Context-aware routing:** The IoV is rich in contextual information, including vehicle positions, speeds, traffic conditions, and environmental factors. Incorporating context awareness into routing algorithms can optimize route selection based on the current context and improve overall network performance. Future research should explore context-aware routing algorithms that leverage contextual information for intelligent and adaptive routing decisions.
- **Multi-hop communication:** In the IoV, vehicles may need to rely on multi-hop communication to reach distant destinations or overcome connectivity gaps. Research should focus on developing efficient multi-hop routing algorithms that can optimize message forwarding and ensure reliable end-to-end communication.
- **Heterogeneous network integration:** The IoV encompasses various communication technologies, such as cellular networks, dedicated short-range communication (DSRC), and Wi-Fi. Integrating these heterogeneous networks poses challenges in terms of routing and seamless handover. Future research should explore routing algorithms that effectively integrate and manage heterogeneous networks to provide uninterrupted connectivity.

- Real-time traffic management: Routing algorithms in the IoV should consider real-time traffic information to make informed routing decisions. Future research should explore integrating real-time traffic data, such as congestion and traffic flow information, into routing algorithms to enable efficient traffic management and avoidance.

## VI. CONCLUSION

The IoV is a transformative technology that enables autonomous driving scenarios by connecting people, intelligent vehicle systems, and cyber-physical systems in urban environments. It has attracted significant commercial and research attention due to advancements in computation and communication technologies, such as edge computing, grid computing, parallel processing, big data analysis, web semantics, and artificial intelligence. The IoV network employs protocols like position, map, and path-based to enable intelligent applications related to road conditions, vehicle preemption, and information services. Maintaining routes in the dynamic IoV network, particularly in homogeneous networks like VANETs, presents challenges due to mobile node behavior. To address this, the IoV network is categorized into 1D, 2D, and 3D scenarios, with Plane-based routing focusing on the latter. However, plane-based routing has limitations in real-world scenarios due to the absence of the third dimension. The IoV network is further divided into homogeneous and heterogeneous networks to optimize vehicle utilization, offering the potential for future development. The paper emphasizes the IoV network's layered architecture, communication, data analysis, and recent challenges, providing valuable insights for researchers and industries. It acknowledges including traditional VANETs and large-scale heterogeneous network structures within the IoV network.

This study has certain limitations that provide opportunities for future research. Firstly, our classification and analysis of IoV routing protocols are primarily based on a theoretical framework. While this provides a structured understanding of these protocols, practical evaluations through extensive simulations and real-world experiments are necessary to validate their performance under varying network conditions. Secondly, the study focuses on existing IoV routing protocols but does not explore the development of novel routing solutions tailored to emerging IoV challenges, such as autonomous vehicles, the integration of 5G, and the proliferation of IoT devices in urban environments. Future research could delve into the design and evaluation of innovative routing protocols that effectively address these evolving demands. Moreover, the security aspects of IoV routing, including privacy preservation and protection against cyberattacks, remain a critical concern. Investigating the integration of robust security measures into routing protocols is essential for ensuring the integrity and confidentiality of data in vehicular networks. Lastly, the study primarily addresses routing at the network layer, and future work could explore the interactions between routing protocols and higher-layer applications, such as traffic management and autonomous vehicle coordination. These areas represent promising avenues for enhancing the efficiency, safety, and overall performance of IoV networks.

## ACKNOWLEDGMENT

This work was supported by the Natural Science Foundation of the Higher Education Institutions of Jiangsu Province (No.22KJD460005).

This work was supported by the Scientific Research Foundation of Nanjing Institute of Technology (No. YKJ201994).

## REFERENCES

- [1] T. Palit, A. M. Bari, and C. L. Karmaker, "An integrated Principal Component Analysis and Interpretive Structural Modeling approach for electric vehicle adoption decisions in sustainable transportation systems," *Decision Analytics Journal*, vol. 4, p. 100119, 2022.
- [2] A. Bhargava, D. Bhargava, P. N. Kumar, G. S. Sajja, and S. Ray, "Industrial IoT and AI implementation in vehicular logistics and supply chain management for vehicle mediated transportation systems," *International Journal of System Assurance Engineering and Management*, vol. 13, no. Suppl 1, pp. 673-680, 2022.
- [3] C. Zhang et al., "Flexible grid-based electrolysis hydrogen production for fuel cell vehicles reduces costs and greenhouse gas emissions," *Applied Energy*, vol. 278, p. 115651, 2020.
- [4] Y. Fu, C. Li, F. R. Yu, T. H. Luan, and Y. Zhang, "A survey of driving safety with sensing, vehicular communications, and artificial intelligence-based collision avoidance," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 6142-6163, 2021.
- [5] B. Pourghebleh and N. Jafari Navimipour, "Towards efficient data collection mechanisms in the vehicular ad hoc networks," *International Journal of Communication Systems*, vol. 32, no. 5, p. e3893, 2019.
- [6] P. Shah and T. Kasbe, "A review on specification evaluation of broadcasting routing protocols in VANET," *Computer Science Review*, vol. 41, p. 100418, 2021.
- [7] M. J. N. Mahi et al., "A review on VANET research: Perspective of recent emerging technologies," *IEEE Access*, 2022.
- [8] Z. Yang, L. Li, F. Gu, and X. Ling, "Dependable and reliable cloud-based architectures for vehicular communications: A systematic literature review," *International Journal of Communication Systems*, vol. 36, no. 7, p. e5457, 2023.
- [9] N. H. Hussein, C. T. Yaw, S. P. Koh, S. K. Tiong, and K. H. Chong, "A Comprehensive Survey on Vehicular Networking: Communications, Applications, Challenges, and Upcoming Research Directions," *IEEE Access*, vol. 10, pp. 86127-86180, 2022.
- [10] J. Webber, A. Mehdodniya, A. Arafat, and A. Alwakeel, "Improved Human Activity Recognition Using Majority Combining of Reduced-Complexity Sensor Branch Classifiers," *Electronics*, vol. 11, no. 3, p. 392, 2022.
- [11] C. Han and X. Fu, "Challenge and Opportunity: Deep Learning-Based Stock Price Prediction by Using Bi-Directional LSTM Model," *Frontiers in Business, Economics and Management*, vol. 8, no. 2, pp. 51-54, 2023.
- [12] R. Soleimani and E. Lobaton, "Enhancing Inference on Physiological and Kinematic Periodic Signals via Phase-Based Interpretability and Multi-Task Learning," *Information*, vol. 13, no. 7, p. 326, 2022.
- [13] S. N. H. Bukhari, J. Webber, and A. Mehdodniya, "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.
- [14] B. M. Jafari, M. Zhao, and A. Jafari, "Rumi: An Intelligent Agent Enhancing Learning Management Systems Using Machine Learning Techniques," *Journal of Software Engineering and Applications*, vol. 15, no. 9, pp. 325-343, 2022.
- [15] J. Webber, A. Mehdodniya, 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.
- [16] H. Kosarirad, M. Ghasempour Nejati, A. Saffari, M. Khishe, and M. Mohammadi, "Feature Selection and Training Multilayer Perceptron Neural Networks Using Grasshopper Optimization Algorithm for Design

- Optimal Classifier of Big Data Sonar," *Journal of Sensors*, vol. 2022, 2022.
- [17] M. Shahin et al., "Cluster-based association rule mining for an intersection accident dataset," in *2021 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube)*, 2021: IEEE, pp. 1-6, doi: 10.1109/ICECube53880.2021.9628206.
- [18] S. Saeidi, S. Enjedani, E. Alvandi Behineh, K. Tehrani, and S. Jazayerifar, "Factors Affecting Public Transportation Use during Pandemic: An Integrated Approach of Technology Acceptance Model and Theory of Planned Behavior," *Tehnički glasnik*, vol. 18, pp. 1-12, 09/01 2023, doi: 10.31803/tg-20230601145322.
- [19] T. Kayarga and S. A. Kumar, "A study on various technologies to solve the routing problem in Internet of Vehicles (IoV)," *Wireless Personal Communications*, vol. 119, pp. 459-487, 2021.
- [20] X. Xu, H. Li, W. Xu, Z. Liu, L. Yao, and F. Dai, "Artificial intelligence for edge service optimization in internet of vehicles: A survey," *Tsinghua Science and Technology*, vol. 27, no. 2, pp. 270-287, 2021.
- [21] R. Gasmı, M. Aliouat, and H. Seba, "A stable link based zone routing protocol (SL-ZRP) for internet of vehicles environment," *Wireless Personal Communications*, vol. 112, pp. 1045-1060, 2020.
- [22] B. Ji et al., "Survey on the internet of vehicles: Network architectures and applications," *IEEE Communications Standards Magazine*, vol. 4, no. 1, pp. 34-41, 2020.
- [23] F. Kamalov, B. Pourghebleh, M. Gheisari, Y. Liu, and S. Moussa, "Internet of Medical Things Privacy and Security: Challenges, Solutions, and Future Trends from a New Perspective," *Sustainability*, vol. 15, no. 4, p. 3317, 2023.
- [24] M. Mohseni, F. Amirghafouri, and B. Pourghebleh, "CEDAR: A cluster-based energy-aware data aggregation routing protocol in the internet of things using capuchin search algorithm and fuzzy logic," *Peer-to-Peer Networking and Applications*, pp. 1-21, 2022.
- [25] 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.
- [26] B. Pourghebleh, V. Hayyolalam, and A. A. Anvigh, "Service discovery in the Internet of Things: review of current trends and research challenges," *Wireless Networks*, vol. 26, no. 7, pp. 5371-5391, 2020.
- [27] B. Pourghebleh, N. Hekmati, Z. Davoudnia, and M. Sadeghi, "A roadmap towards energy-efficient data fusion methods in the Internet of Things," *Concurrency and Computation: Practice and Experience*, p. e6959, 2022.
- [28] S. El Madani, S. Motahhir, and A. El Ghzizal, "Internet of vehicles: concept, process, security aspects and solutions," *Multimedia Tools and Applications*, vol. 81, no. 12, pp. 16563-16587, 2022.
- [29] A. Hbaieb, S. Ayed, and L. Chaari, "A survey of trust management in the Internet of Vehicles," *Computer Networks*, vol. 203, p. 108558, 2022.
- [30] N. K. Narang, "Mentor's Musings on the Role of Standards in Improving the Privacy, Trust and Reputation Management in Internet of Vehicles (IoV)," *IEEE Internet of Things Magazine*, vol. 6, no. 2, pp. 18-23, 2023.
- [31] A. Salim, A. M. Khedr, B. Alwasel, W. Osamy, and A. Aziz, "SOMACA: A New Swarm Optimization-Based and Mobility-Aware Clustering Approach for the Internet of Vehicles," *IEEE Access*, 2023.
- [32] A. Guerna, S. Bitam, and C. T. Calafate, "Roadside unit deployment in internet of vehicles systems: A survey," *Sensors*, vol. 22, no. 9, p. 3190, 2022.
- [33] N. Azzaoui, A. Korichi, B. Brik, and H. Amirat, "A survey on data dissemination in internet of vehicles networks," *Journal of Location Based Services*, pp. 1-44, 2022.
- [34] B. Ji et al., "A vision of IoV in 5G HetNets: architecture, key technologies, applications, challenges, and trends," *IEEE Network*, vol. 36, no. 2, pp. 153-161, 2022.
- [35] B. P. Rimal, C. Kong, B. Poudel, Y. Wang, and P. Shahi, "Smart electric vehicle charging in the era of internet of vehicles, emerging trends, and open issues," *Energies*, vol. 15, no. 5, p. 1908, 2022.
- [36] G. Husnain, S. Anwar, and F. Shahzad, "An Enhanced AI-Enabled Routing Optimization Algorithm for Internet of Vehicles (IoV)," *Wireless Personal Communications*, vol. 130, no. 4, pp. 2623-2643, 2023.
- [37] R. Dhanare, K. K. Nagwanshi, and S. Varma, "A study to enhance the route optimization algorithm for the internet of vehicle," *Wireless Communications and Mobile Computing*, vol. 2022, 2022.
- [38] N. M. Elfatih et al., "Internet of vehicle's resource management in 5G networks using AI technologies: Current status and trends," *IET Communications*, vol. 16, no. 5, pp. 400-420, 2022.
- [39] S. J. Kamble and M. R. Kounte, "Enabling technologies for internet of vehicles," in *Proceeding of the International Conference on Computer Networks, Big Data and IoT (ICCCBI-2018)*, 2020: Springer, pp. 257-268.
- [40] S. Sharma and B. Kaushik, "Applications and challenges in internet of vehicles: A survey," *Internet of Things and Its Applications: Select Proceedings of ICIA 2020*, pp. 55-65, 2022.
- [41] S. Sharma and B. Kaushik, "A survey on nature-inspired algorithms and its applications in the Internet of Vehicles," *International Journal of Communication Systems*, vol. 34, no. 12, p. e4895, 2021.
- [42] S. Kumar and J. Singh, "Internet of Vehicles over VANETs: smart and secure communication using IoT," *Scalable Computing: Practice and Experience*, vol. 21, no. 3, pp. 425-440, 2020.
- [43] M. Lee and T. Atkison, "Vanet applications: Past, present, and future," *Vehicular Communications*, vol. 28, p. 100310, 2021.
- [44] C. R. Storck and F. Duarte-Figueiredo, "A survey of 5G technology evolution, standards, and infrastructure associated with vehicle-to-everything communications by internet of vehicles," *IEEE access*, vol. 8, pp. 117593-117614, 2020.
- [45] H. Vasudev, D. Das, and A. V. Vasilakos, "Secure message propagation protocols for IoVs communication components," *Computers & Electrical Engineering*, vol. 82, p. 106555, 2020.
- [46] H. H. Jeong, Y. C. Shen, J. P. Jeong, and T. T. Oh, "A comprehensive survey on vehicular networking for safe and efficient driving in smart transportation: A focus on systems, protocols, and applications," *Vehicular Communications*, vol. 31, p. 100349, 2021.
- [47] W. Xu, Y. Zhang, F. Wang, Z. Qin, C. Liu, and P. Zhang, "Semantic Communication for the Internet of Vehicles: A Multiuser Cooperative Approach," *IEEE Vehicular Technology Magazine*, 2023.
- [48] E. Qafzezi, K. Bylykbashi, P. Ampirit, M. Ikeda, K. Matsuo, and L. Barolli, "An intelligent approach for cloud-fog-edge computing SDN-VANETs based on fuzzy logic: effect of different parameters on coordination and management of resources," *Sensors*, vol. 22, no. 3, p. 878, 2022.
- [49] K. N. Qureshi, S. Din, G. Jeon, and F. Piccialli, "Internet of vehicles: Key technologies, network model, solutions and challenges with future aspects," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 3, pp. 1777-1786, 2020.
- [50] D. T. Le, K. Q. Dang, Q. L. T. Nguyen, S. Alhelaly, and A. Muthanna, "A behavior-based malware spreading model for vehicle-to-vehicle communications in VANET networks," *Electronics*, vol. 10, no. 19, p. 2403, 2021.
- [51] A. M. Khasawneh et al., "Service-centric heterogeneous vehicular network modeling for connected traffic environments," *Sensors*, vol. 22, no. 3, p. 1247, 2022.

# Design and Implementation Submarine Cable Object Detection YOLOv4 based with Graphical User Interface (GUI) for Remotely Operated Vehicle (ROV)

Fikri Arif Wicaksana, Eueung Mulyana, Syarif Hidayat, Rahadian Yusuf  
School of Electrical Engineering and Informatics, Institut Teknologi Bandung, Bandung, Indonesia

**Abstract**—The use of submarine cables as underwater transmission channels for distributing electrical energy in Indonesian waters is crucial. However, the detection and maintenance of submarine cables still heavily rely on human observation, leading to limitations in time and subjective interpretations. This research aims to design and implement an underwater object detection system based on YOLOv4 integrated with a Graphical User Interface (GUI) on a Remotely Operated Vehicle (ROV) for submarine cable detection. The YOLOv4 model was trained using a balanced dataset, achieving performance with precision of 0.89, recall of 0.85, and f1-score of 0.87. Detection of Good Condition (SC-Good-Condition) achieved an Average Precision (AP) of 97.62%, while Bad Condition detection (SC-Bad-Condition) had an AP of 87.54%, resulting in an overall mAP of 92.58%. The implemented GUI successfully detected submarine cables in two test videos with FPS rates of 0.178 and 0.083. The designed underwater object detection system using YOLOv4 and GUI on ROV demonstrated satisfactory performance in detecting submarine cables. However, further efforts are needed to improve the GUI's FPS to make it more responsive and efficient. This research contributes to the development of underwater detection technology that supports environmental observation and electrical energy distribution in Indonesian waters.

**Keywords**—Submarine cable; object detection; GUI; ROV; YOLOv4

## I. INTRODUCTION

Electricity is a fundamental and crucial necessity for the livelihood of Indonesian society. As Indonesia's economic growth and population continue to expand, the demand for electricity increases. Furthermore, being an archipelagic nation with 17,504 islands, almost all activities in both rural and urban areas of Indonesia require electricity. Therefore, there is a need for a transmission medium that can distribute electricity from one island to another. Submarine cables are underwater transmission channels that can distribute electricity between Indonesian islands [1]. However, in practice, submarine cables require periodic maintenance to ensure their optimal condition. One stage in the process of maintaining submarine cables is underwater inspection to observe the surrounding environment of the cables.

A Remotely Operated Vehicle (ROV) is an underwater robot that can be controlled remotely [2]. Some ROVs are

equipped with computer vision technology-enabled cameras for underwater inspection purposes, including maintenance and observation in the vicinity of submarine cables [3]. Typically, the process of detecting submarine cables is visually performed by ROV operators who oversee the video feed from the camera mounted on the ROV. However, this approach has some limitations, such as dependency on human observation, which is susceptible to fatigue and time constraints, as well as subjective interpretation in identifying submarine cables [4].

To address these limitations, automated object detection techniques have been developed, including underwater object detection. One method that has shown good performance is YOLOv4 (You Only Look Once version 4), which enables fast and accurate object detection [5]-[7].

Additionally, implementing a Graphical User Interface (GUI) on the ROV can provide an intuitive and user-friendly interface for operators, facilitating cable observation and detection with higher efficiency [8]. However, despite several studies on underwater object detection using YOLOv4 and research on GUI implementation on ROVs, research specifically combining both aspects in the context of submarine cable detection is still limited.

Therefore, this research aims to fill this knowledge gap by designing and implementing an underwater object detection system based on YOLOv4, integrated with a GUI on the ROV, particularly in the context of submarine cable detection. This study is expected to enhance the effectiveness and efficiency of submarine cable detection more accurately and efficiently.

Several prior studies have explored submarine cable detection using various methods, including both Deep Learning and non-Deep Learning approaches, yielding reasonably accurate results.

One previous study employed CNN and YOLO model (YOLOv3) for submarine cable detection but lacked a desktop GUI application. The results were as follows: for the original image dataset, it achieved an Average Precision (AP) of 98.14%, an F1 Score of 95.79%, and an Average Time of 0.416 seconds. Meanwhile, after dataset enhancement, it achieved an AP of 98.95%, an F1 Score of 96.92%, and an Average Time of 0.452 seconds [9].

Another research used edge detection methods to detect submarine cables, utilizing the Hough Transform model. This study developed software for ROVs equipped with cameras for cable detection. It reported successful submarine cable detection using 100 scenes, resulting in 83 correct detections, 17 false detections, and a recognition rate of 83% in the first experiment. In the second experiment, out of 100 scenes, it achieved 98 correct detections, 2 false detections, and a recognition rate of 98% [10][11].

This research focuses on submarine cable detection using the more recent CNN model YOLOv4, enhanced by the inclusion of a GUI for computer vision on the ROV. This integration of YOLOv4 and GUI aims to improve the quality of ROV-acquired computer vision data compared to previous studies, providing a more advanced and comprehensive approach to submarine cable detection.

In line with the outlined objectives, this paper is organized as follows. Section I begins with an exploration of the background of the research, emphasizing the vital role of electricity in Indonesian nation, the challenges posed by its archipelagic nature, and the need for advanced techniques in submarine cable detection. It further elucidates the specific goals and contributions of this research, aiming to bridge the existing gap in the integration of YOLOv4-based object detection and GUI on ROVs for submarine cable detection. Moreover, this section provides a concise review of prior research endeavors related to underwater object detection, ROV applications, and GUI implementations in the marine domain.

Section II delves into the research design, detailing the methodologies, equipment, and procedures employed in developing the underwater object detection system using YOLOv4 and integrating it with the ROV's GUI. It sheds light on the technical aspects and considerations pivotal to the success of this innovative system.

Section III is dedicated to presenting the experimental findings and their subsequent analysis. The section elucidates the outcomes of deploying the YOLOv4-based system and GUI on the ROV during underwater cable inspections. It discusses performance metrics and GUI functionality test.

Section IV, we synthesize the findings into comprehensive conclusions and offer recommendations for future research in this field. We reflect on the implications of our work on the broader context of submarine cable maintenance and its potential contributions to the sustainability and efficiency of Indonesia's electricity distribution network. Furthermore, we suggest avenues for further research and improvements to enhance the capabilities of the integrated system, ensuring its continued effectiveness in the evolving landscape of underwater cable detection.

## II. RESEARCH DESIGN

### A. Research Stages

There are several steps carried out in this research, as shown in Figure 1.

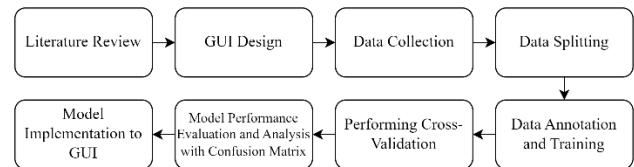


Fig. 1. Research stages.

- **Literature Review:** Conducting a comprehensive literature review on underwater object detection methods and techniques, the YOLOv4 approach, GUI utilization on ROVs, and related submarine cable research to understand the foundational theories and related studies.
- **GUI Design:** Designing and developing the Graphical User Interface (GUI) using PyQt (Python QML) library to enable interaction between the ROV operator and the YOLOv4-based underwater object detection system.
- **Data Collection:** Gathering the necessary image data for training and testing the submarine cable detection model. The image data used in this research was obtained from underwater inspection videos conducted by PT Syergie Indoprima in the year 2022.
- **Data Splitting:** Dividing the image data into two classes: SC-Good-Condition (good cable condition) and SC-Bad-Condition (bad cable condition). The entire image data will be divided into two subsets, with 80% of the data used for model training and 20% for model testing, both for imbalanced and balanced datasets.
- **Data Annotation and Training:** Annotating the image data with appropriate labels, i.e., SC-Good-Condition or SC-Bad-Condition, indicating the condition of the submarine cable in each image. Next, training the underwater object detection model using YOLOv4 with the annotated image data.
- **Performing Cross-Validation:** Conducting cross-validation on the trained model to measure its performance and accuracy in detecting submarine cables. The data is divided into 5 subsets (folds), and the model training and testing will be repeated on each fold to obtain more reliable results.
- **Model Performance Evaluation and Analysis with Confusion Matrix:** Using the cross-validation results, a confusion matrix is created to depict the overall model performance. The confusion matrix provides information on the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) in detecting submarine cables. Through the confusion matrix, the model's performance is analyzed and evaluated, including calculation of evaluation metrics such as precision, recall, and F1-score, to gain a deeper understanding of the model's ability to detect submarine cable conditions.

- **Model Implementation into GUI:** The trained and evaluated underwater object detection model is then implemented into the previously designed GUI application. This process involves a seamless integration between the model and GUI to enable real-time and non-real-time submarine cable detection through the ROV using a user-friendly GUI that provides clear and understandable results for the ROV operator.

**B. GUI Design**

Figure 2 is a mock-up of the GUI designed in this research. This GUI functions to display real-time and non-real-time computer vision camera views for detecting submarine cables.

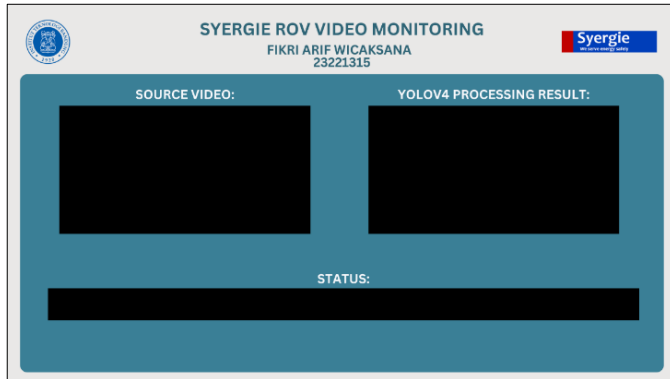


Fig. 2. Mock-up of ROV computer vision application GUI.

This GUI is equipped with 2 screen views, namely the "source video" screen to display real-time and non-real-time computer vision and the "YOLOv4 Processing Result" screen to show the submarine cable detection results from the trained YOLOv4 weights. Additionally, there is a status bar that displays the GUI duration, submarine cable detection class, AP (Average Precision) value, and bounding box dimensions.

**C. Data Collection**

The data used in this study was collected from underwater inspection video documentation by PT Syergie Indoprima. The documentation videos were converted into image frames using VLC Media Player. There are 4 underwater inspection videos used in this research, resulting in a total of 395 image frames. The data collection process is illustrated in Figure 3.

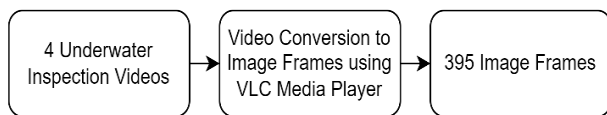


Fig. 3. Collecting data stages.

**D. Data Splitting**

After the data collection process, the data is divided into two classes: SC-Good-Condition and SC-Bad-Condition. This class division is based on the physical conditions observed in the underwater inspection videos that have been converted into 395 image frames. The following are some criteria for the submarine cable classes in this research, presented in Table I.

TABLE I. CRITERIA FOR SUBMARINE CABLE CLASSES

Criteria for SC-Good-Condition	Criteria for SC-Bad-Condition
The armor layer of the submarine cable is intact, with no peeling.	The armor layer of the submarine cable is peeling.
The submarine cable is not covered by underwater vegetation.	The submarine cable is covered by underwater vegetation [12]

Figure 4 is an example of SC-Good-Condition image, and Figure 5 is an example of SC-Bad-Condition image.



Fig. 4. Example of SC-good-condition image.

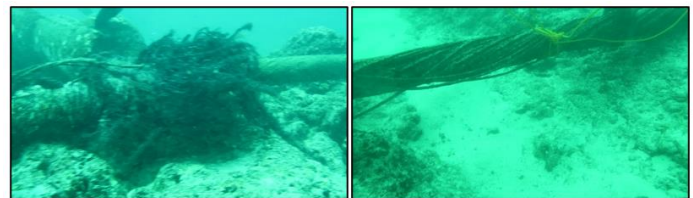


Fig. 5. Example of SC-Bad-condition image.

Out of the 395 submarine cable image frames obtained, the number of images for each class was imbalanced. Therefore, the image data was re-divided to create a balanced dataset for the research. The balanced dataset used is presented in Table II, and the data split between the training and testing sets is presented in Table III.

TABLE II. DATA IMAGE SPLIT INTO BALANCED DATASET

Class	Number of Images
SC-Good-Condition	100 Images
SC-Bad-Condition	100 Images

TABLE III. DATA IMAGE SPLIT BETWEEN TRAINING AND TESTING SETS

Number of Image Data	Training Data (80%)	Test Data (20%)
200 Images	160 Images	40 Images

Balanced and imbalanced datasets are important concepts in Deep Learning, including for object detection models like YOLOv4<sup>1</sup>. When collecting image data, it is crucial to pay

<sup>1</sup>Introduction to Balanced and Imbalanced Datasets in Machine Learning. (n.d.). Balanced and Imbalanced Datasets in Machine Learning [Full Introduction]. <https://encord.com/blog/an-introduction-to-balanced-and-imbalanced-datasets-in-machine-learning/>

attention to dataset balance, especially if there are differences in the number of examples between the SC-Good-Condition and SC-Bad-Condition classes. If the dataset is imbalanced, for instance, when there are fewer examples of bad submarine cables compared to good examples, the object detection model can become biased in learning relevant patterns and features of the bad submarine cables [13]. Therefore, it is necessary to perform proportional data splitting for training and testing the model effectively, enabling the model to accurately recognize both classes and avoid any undesired bias in submarine cable detection.

E. Data Annotation and Training

In this research, the tool LabelIm<sup>2</sup> was utilized for annotating the images used in submarine cable detection. This tool allows users to easily create annotations in the YOLOv4 format. It provides features to manually select and mark the positions and boundaries of submarine cables for both SC-Good-Condition and SC-Bad-Condition classes. The annotation process involves drawing bounding boxes around the cables in each image. These bounding boxes provide information about the coordinates (x, y) and dimensions (width and height) of the submarine cables.

For the training data process, custom configurations were used based on the number of classes being trained. A reference guide<sup>3</sup> was employed for the configuration of submarine cable detection, as presented in Table IV for yolov4\_train.cfg and Table V for yolov4\_test.cfg.

TABLE IV. CONFIGURATION FOR YOLOV4\_TRAIN.CFG

Reference	Used Configuration
$Filters = (number\ of\ class + 5) \times B$ B = number of bounding boxes	$Filters = (2 + 5) \times 3$ $Filters = 21$
$Max\ batches =$ $number\ of\ classes \times 2000$	$Max\ batches = 2 \times 2000$ $Max\ batches = 4000$
$Steps = 80\%, 90\% \text{ of } max\ batches$	$Steps = 80\%, 90\% \text{ of } 4000$ $Steps = 3200, 3600$
$Batch = 32$ $Subdivision = 16$	$Batch = 32$ $Subdivision = 16$

TABLE V. CONFIGURATION FOR YOLOV4\_TEST.CFG

Reference [14]	Used Configuration
$Filters = (number\ of\ class + 5) \times B$ B = number of bounding boxes	$Filters = (2 + 5) \times 3$ $Filters = 21$
$Max\ batches =$ $number\ of\ classes \times 2000$	$Max\ batches = 2 \times 2000$ $Max\ batches = 4000$
$Steps = 80\%, 90\% \text{ of } max\ batches$	$Steps = 80\%, 90\% \text{ of } 4000$ $Steps = 3200, 3600$
$Batch = 1$ $Subdivision = 1$	$Batch = 1$ $Subdivision = 1$

<sup>2</sup>H. (2022, September 22). GitHub - heartexlabs/labelImg: LabelImg is now part of the Label Studio community. The popular image annotation tool created by Tzutalin is no longer actively being developed, but you can check out Label Studio, the open source data labeling tool for images, text, hypertext, audio, video and time-series data. GitHub. <https://github.com/heartexlabs/labelImg>

<sup>3</sup>A. (2023, June 20). GitHub - AlexeyAB/darknet: YOLOv4 / Scaled-YOLOv4 / YOLO - Neural Networks for Object Detection (Windows and Linux version of Darknet ). GitHub. <https://github.com/AlexeyAB/darknet>

- Filters: In the YOLOv4 configuration (cfg) file, the "filters" parameter refers to the number of filters needed in the last convolutional layer of the YOLOv4 network architecture. This number of filters is related to the number of object classes to be detected (plus 5), then multiplied by 3.  $Filters = (Number\ of\ classes + 5) \times B$  indicates that for each object class to be detected, plus 5 (representing bounding box coordinates and confidence score), it will be multiplied by 3. The result is the total number of filters required in the last convolutional layer.
- Max batches: In the YOLOv4 configuration (cfg) file, the "max\_batches" parameter determines the total number of iterations that will be used in the model training. Each iteration involves one batch of image data to train the model. The value of "max\_batches" indicates the limit on the number of iterations to be performed during training. Therefore, from the formula  $Max\ batches = Number\ of\ classes \times 2000$ , it can be concluded that in this research, the maximum number of iterations for each training session is set to 4000 iterations.
- Steps: In the YOLOv4 configuration (cfg) file, the "steps" parameter is used to control when the learning rate will be adjusted during the training process. The "steps" value, expressed as a percentage of "max\_batches," determines the points at which the learning rate will undergo changes. In this research, "steps" are set at 80% and 90% of "max\_batches," which means there are two points where the learning rate will change during training: (1) At 80% of "max\_batches": The learning rate will be adjusted when the training reaches 80% of the total scheduled iterations ("max\_batches"). This adjustment usually involves decreasing the learning rate to help the model reach an optimal point during training. (2) At 90% of "max\_batches": The learning rate will be adjusted when the training reaches 90% of the total scheduled iterations ("max\_batches"). This adjustment typically involves further decreasing the learning rate to smooth the model's convergence process and improve the final results. The purpose of adjusting the learning rate is to optimize the training process and aid the model in achieving a good convergence, thereby enhancing object detection performance.
- Batch and subdivision: In the YOLOv4 configuration (cfg) file, the "batch" and "subdivisions" parameters are used to control how training or testing data is processed in each iteration. Here is an explanation for both values: (1) For the "yolov4\_train.cfg":  $Batch=32$ : The "batch" value indicates the number of images processed in each training iteration. In this case, 32 images are processed simultaneously in one iteration. This means that 32 images are loaded into memory and used to update the model weights in one iteration.  $Subdivisions=16$ : The "subdivisions" value indicates how many weight updates will be performed before considering one iteration complete. In this case, every 16 weight updates will be performed before one

iteration is considered complete. This is useful to reduce memory load and speed up the training process. With this configuration, in each training iteration, 32 images are loaded and processed, and weight updates are performed every 16 times. This allows for YOLOv4 model training with efficiency and speeds up the training process. (2) For the "yolov4\_test.cfg": Batch=1: The "batch" value is set to 1, which means that in each testing iteration, only 1 image will be processed. This is done because during the testing phase, we want to test each image individually to obtain accurate detection results and precise evaluation. Subdivisions=1: The "subdivisions" value is set to 1, which means that weight updates are not needed during the testing process. Each image is tested separately without any weight updates because no training is done at this stage. With this configuration, each image in the testing data will be tested separately, one by one, without any weight updates performed. This allows for accurate testing and precise evaluation of the previously trained model.

#### F. Performing Cross-Validation

In this research, cross-validation method is used to evaluate and validate the performance of the underwater object detection model. Cross-validation is a statistical method employed to assess and validate a model on a limited dataset<sup>4</sup>. The following are the steps of cross-validation conducted in this study:

- **Data Splitting:** The dataset used is divided into several subsets called "folds". In this research, the dataset is divided into 5 folds, as shown in the data split in Table VI.
- **Cross-Validation Iterations:** The cross-validation is performed 5 times, where each iteration uses one fold as the testing data, and the remaining folds are used as the training data. For example, in the first iteration, fold 5 is used as the testing data, while folds 1 to 4 are used as the training data. In the second iteration, fold 4 is used as the testing data, and folds 5 and 1 to 3 are used as the training data, and so on.
- **Model Training:** In each iteration, the submarine cable detection model is trained using the designated training data. The training process is conducted using the predetermined techniques and parameters.
- **Model Testing:** After training the model in each iteration, the model is evaluated using separate testing data. The model's performance is measured using relevant evaluation metrics such as precision, recall, and F1-score.
- **Selecting the Best Iteration Result:** After completing the cross-validation iterations, the performance evaluation results of the model in each iteration are recorded. Then, the best iteration result is chosen based on evaluation metrics like precision, recall, and F1-

score. The best iteration result is selected as the final outcome representing the model's best performance.

TABLE VI. DATASET FOLD SPLITTING

Iteration	Fold 1 (20%)	Fold 2 (20%)	Fold 3 (20%)	Fold 4 (20%)	Fold 5 (20%)
Iteration 1	Train				Test
Iteration 2	Train			Test	Train
Iteration 3	Train		Test	Train	
Iteration 4	Train	Test	Train		
Iteration 5	Test	Train			

Through the cross-validation method, this research can obtain more stable and reliable estimations of the performance of the submarine cable detection model. By dividing the dataset into different subsets for training and testing, this study can objectively test the model on various data and identify its strengths and weaknesses.

#### G. Model Performance Evaluation and Analysis with Confusion Matrix

This stage involves using the confusion matrix to evaluate and analyze the performance of the submarine cable detection model. The following are the steps involved:

- **Building the Confusion Matrix:** Using the testing data, the submarine cable detection model will make predictions for the SC-Good-Condition and SC-Bad-Condition object classes. From the prediction results and the ground truth labels, the confusion matrix will be constructed. The confusion matrix is a table with four cells representing the number of correct and incorrect predictions for each object class. The cells in the confusion matrix include True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) [14]. The visualization of the confusion matrix is shown in Figure 6.
- **Based on the visualization of the confusion matrix above, in the context of submarine cable detection with classes SC-Good-Condition and SC-Bad-Condition, the definitions of True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) can be stated as follows:**
  - **True Positive (TP):** TP occurs when the model correctly detects a submarine cable in good condition (SC-Good-Condition). This means that the model predicts correctly that the example belongs to the SC-Good-Condition class and indeed represents a submarine cable in good condition.
  - **True Negative (TN):** TN occurs when the model correctly detects a submarine cable in bad condition (SC-Bad-Condition). This means that the model predicts correctly that the example belongs to the SC-Bad-Condition class and indeed represents a submarine cable in bad condition.

<sup>4</sup>Cross Validation: Teknik Evaluasi Machine Learning, 6 Metode. (2022, August 14). Digital Polar. <https://digitalpolar.com/cross-validation/>



- False Positive (FP): FP occurs when the model incorrectly detects a submarine cable in bad condition (SC-Bad-Condition) as a submarine cable in good condition (SC-Good-Condition). This means that the model mistakenly predicts that the example belongs to the SC-Good-Condition class, whereas it actually belongs to the SC-Bad-Condition class.
- False Negative (FN): FN occurs when the model incorrectly detects a submarine cable in good condition (SC-Good-Condition) as a submarine cable in bad condition (SC-Bad-Condition). This means that the model mistakenly predicts that the example belongs to the SC-Bad-Condition class, whereas it actually belongs to the SC-Good-Condition class.

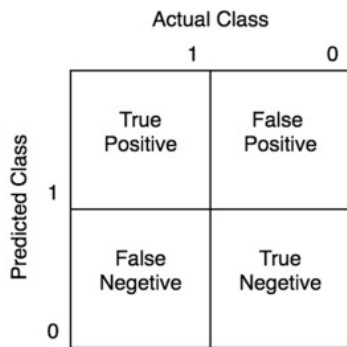


Fig. 6. Confusion matrix visualization [15].

- Calculating Evaluation Metrics: Based on the confusion matrix, various performance evaluation metrics for the model can be calculated, such as precision, recall, and F1-score. Precision measures the extent to which the model's positive predictions are correct, while recall measures the extent to which the model can correctly identify positive objects. F1-score is a combined measure that takes into account both precision and recall to provide a balanced performance overview<sup>5</sup>. The calculation algorithms for each metric are presented in Table VII.

TABLE VII. ALGORITHM FOR CALCULATING EVALUATION METRICS

Metric	Calculation Algorithm
Precision	$\frac{TP}{(TP + FP)}$ (1)
Recall	$\frac{TP}{TP + FN}$ (2)
F1-score	$\frac{2 \times (Precision \times Recall)}{(Precision + Recall)}$ (3)
mAP	$\frac{1}{N} \sum_{i=1}^N AP_i$ (4) Note: N = Number of AP data AP = Average Precision

<sup>5</sup>Kumar, A. (2023, March 17). Accuracy, Precision, Recall & F1-score - Python Examples - Data Analytics. Data Analytics. <https://vitalflux.com/accuracy-precision-recall-f1-score-python-example>

- Interpretation of Results: Through the confusion matrix and calculated evaluation metrics, the performance results of the model can be interpreted. Analyzing the number of TP, TN, FP, and FN for each object class provides insights into the model's ability to detect submarine cables in both SC-Good-Condition and SC-Bad-Condition classes. Observing the values of precision, recall, and F1-score for each object class helps in understanding the strengths and weaknesses of the model in detecting submarine cables.
- Visualization of Confusion Matrix: To facilitate understanding, the confusion matrix can be visualized using graphs or heat maps. This helps to clearly visualize the distribution of prediction results and errors that occur across all object classes.
- Testing Submarine Cable Detection on 20 Image Samples: To ensure that the model can accurately detect submarine cable conditions, a test for submarine cable detection is performed on 20 images listed in Table VIII, which are then analyzed for their results.

TABLE VIII. DETAILS OF THE SUBMARINE CABLE DETECTION TEST IMAGES

File Name	Class
Sample01.png – Sample10.png	SC-Good-Condition
Sample11.png – Sample20.png	SC-Bad-Condition

Through the evaluation and performance analysis of the model using the confusion matrix, this research provides detailed insights into how the submarine cable detection model operates, the extent of errors that occur, and the model's performance across all object classes. This aids in understanding and reporting the model's performance more accurately and informatively.

#### H. Model Implementation to GUI

This stage involves integrating the object detection model for underwater objects into the previously designed GUI. The following are the steps involved in this implementation:

- Model Preparation: The trained and tested submarine cable detection model (weights) will be prepared for integration into the GUI.
- Integration with GUI Library and Framework: The model will be integrated with the GUI library and framework used in this research, which is PyQt (Python QML).
- Functional Testing: After the integration is complete, functional testing of the GUI will be conducted to ensure that the submarine cable detection model operates smoothly within the GUI. At this stage, non-real-time submarine cable detection will be tested on several videos, and the details of these test videos are presented in Table IX.
- Evaluation and Refinement: After testing the functionality, the performance of the GUI and the submarine cable detection model within the GUI

environment is evaluated. Several aspects that can be evaluated include the mAP of the detection results according to equation (4) and also the calculation of the GUI's Frames per Second (FPS) performance, which can be calculated using the following equation:

$$FPS = \frac{\text{Number of Detected SC Frames}}{\text{GUI Processing Time}} \quad (5)$$

TABLE IX. DETAIL OF SUBMARINE CABLE DETECTION TEST VIDEOS ON GUI.

File Name	Video Duration
SampelVideo1.mov	54 seconds
SampleVideo2.mov	60 seconds

Subsequently, areas that need improvement or enhancement can be identified, and refinements can be made as necessary.

By integrating the submarine cable detection model into the GUI, this research provides an intuitive and interactive interface for users to perform submarine cable detection practically and efficiently. The GUI facilitates both real-time and non-real-time usage of the model and streamlines the interpretation of detection results within a more structured environment.

### III. RESULT AND ANALYSIS

In this chapter, the results of the experiments, evaluation, and performance analysis of the model, as well as the functionality test of the GUI, are explained in accordance with the previously described design.

#### A. Experiment using Imbalanced Dataset

A total of 16 training experiments were conducted using the imbalanced dataset, wherein four datasets with varying numbers of images were used. The number of images in the SC-Good-Condition class was smaller compared to the number of images in the SC-Bad-Condition class, with a ratio of 40:60. The details of the image distribution for the imbalanced dataset are presented in Table X and Table XI.

From the division of the imbalanced dataset, training was conducted with varying numbers of iterations. Each dataset was trained using 1000, 2000, 3000, and 4000 iterations. The training results produced 16 metric outcomes, presented in Table XII and visualized in the graph in Figure 7.

TABLE X. THE NUMBER OF IMAGES IN THE SC-GOOD-CONDITION CLASS AND THE SC-BAD-CONDITION CLASS USED IN THE IMBALANCED DATASET

Dataset Name	Number of Images Data	SC-Good-Condition Class (40%)	SC-Bad-Condition Class (60%)
Dataset 1	395 Images	158 Images	237 Images
Dataset 2	790 Images	316 Images	474 Images
Dataset 3	1185 Images	474 Images	711 Images
Dataset 4	1580 Images	632 Images	948 Images

TABLE XI. COMPOSITION OF IMBALANCED DATASET

Dataset Name	Original Images	Rotated Images 90°	Rotated Images 180°	Rotated Images 270°	Total Number of Image Data
Dataset 1	395 Images	0 Images	0 Images	0 Images	395 Images
Dataset 2	395 Images	395 Images	0 Images	0 Images	790 Images
Dataset 3	395 Images	395 Images	395 Images	0 Images	1185 Images
Dataset 4	395 Images	395 Images	395 Images	395 Images	1580 Images

TABLE XII. TRAINING RESULTS USING THE IMBALANCED DATASET

Name of Weight	Precision	Recall	F1-score	mAP @0.5
yolov4_imb_395_1000	0.69	0.58	0.63	64.62%
yolov4_imb_395_2000	0.87	0.82	0.85	87.67%
yolov4_imb_395_3000	0.87	0.83	0.85	87.37%
yolov4_imb_395_4000	0.87	0.83	0.85	87.67%
yolov4_imb_790_1000	0.71	0.14	0.24	45.04%
yolov4_imb_790_2000	0.72	0.59	0.64	49.95%
yolov4_imb_790_3000	0.76	0.65	0.70	68.58%
yolov4_imb_790_4000	0.82	0.64	0.72	70.98%
yolov4_imb_1185_1000	0.62	0.06	0.11	18.86%
yolov4_imb_1185_2000	0.70	0.26	0.38	33.15%
yolov4_imb_1185_3000	0.70	0.24	0.36	31.84%
yolov4_imb_1185_4000	0.69	0.23	0.35	30.97%
yolov4_imb_1580_1000	0.54	0.05	0.09	17.18%
yolov4_imb_1580_2000	0.65	0.29	0.40	37.52%
yolov4_imb_1580_3000	0.62	0.29	0.40	38.12%
yolov4_imb_1580_4000	0.69	0.27	0.38	38.63%

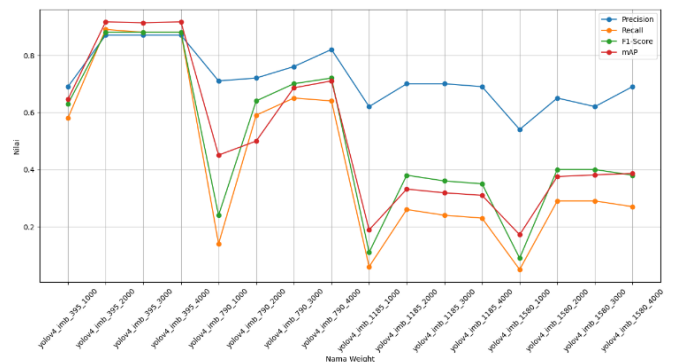


Fig. 7. Graph of evaluation metrics from training using the imbalanced dataset.

From Table XII and Figure 7, it can be observed that there is variation in the performance metrics across different datasets and iterations. It is evident that increasing the number of images in the dataset and the number of iterations does not always result in consistent improvement in the measured metrics.

For example, in the mAP (mean Average Precision) column, it can be seen that some combinations have lower

values compared to others, even when using a higher number of images and iterations. For instance, in the "yolov4\_imb\_395" dataset, the mAP for 4000 iterations (87.67%) does not show a significant improvement compared to 2000 iterations (87.67%). This indicates that after reaching a certain threshold of images and iterations, further increases do not provide significant benefits in the measured metrics.

This highlights that performance improvement is not solely dependent on increasing the number of data and iterations. There are other factors that need to be considered, such as data quality, balance of data between classes, and algorithmic factors used in model training. In this experiment with the imbalanced dataset, the best metric results were achieved by the "yolov4\_imb\_395\_4000" weight.

**B. Experiment using Balanced Dataset**

A total of 20 training experiments were conducted with the balanced dataset using the cross-validation method. Among these experiments, the training process involved 5 iterations of cross-validation, with 5 fold data images for both training and testing data, as indicated in Table VI. Each cross-validation iteration utilized a different number of training iterations, similar to the training of the imbalanced dataset, which included 1000, 2000, 3000, and 4000 training iterations.

The number of images in the SC-Good-Condition class was balanced with the number of images in the SC-Bad-Condition class, with a 1:1 ratio. The details of image distribution for the balanced dataset are provided in Table II and Table III, as discussed previously. The training results from the balanced dataset yielded 20 sets of performance metrics, presented in Table XIII, and visualized in Figure 8.

TABLE XIII. TRAINING RESULTS USING THE BALANCED DATASET

Name of Weight	Precision	Recall	F1-score	mAP @0.5
yolov4_blc_itr1_1000	0.62	0.32	0.42	51.98%
yolov4_blc_itr1_2000	0.70	0.68	0.69	64.54%
yolov4_blc_itr1_3000	0.80	0.78	0.79	75.65%
yolov4_blc_itr1_4000	0.79	0.76	0.77	72.86%
yolov4_blc_itr2_1000	0.78	0.62	0.69	73.68%
yolov4_blc_itr2_2000	0.86	0.80	0.83	86.06%
yolov4_blc_itr2_3000	0.81	0.73	0.76	78.44%
yolov4_blc_itr2_4000	0.89	0.85	0.87	92.58%
yolov4_blc_itr3_1000	0.55	0.45	0.49	51.28%
yolov4_blc_itr3_2000	0.91	0.77	0.84	77.01%
yolov4_blc_itr3_3000	0.84	0.77	0.81	77.08%
yolov4_blc_itr3_4000	0.89	0.77	0.83	77.53%
yolov4_blc_itr4_1000	0.55	0.38	0.45	36.53%
yolov4_blc_itr4_2000	0.70	0.55	0.61	49.87%
yolov4_blc_itr4_3000	0.53	0.38	0.44	34.20%
yolov4_blc_itr4_4000	0.75	0.57	0.65	61.09%
yolov4_blc_itr5_1000	0.47	0.35	0.40	40.90%
yolov4_blc_itr5_2000	0.59	0.47	0.53	47.13%
yolov4_blc_itr5_3000	0.72	0.57	0.64	59.87%
yolov4_blc_itr5_4000	0.82	0.68	0.74	72.25%

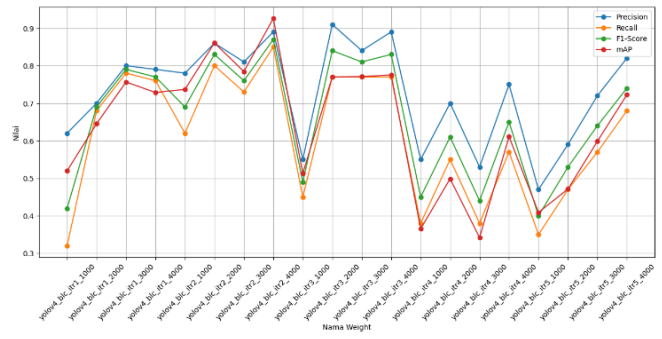


Fig. 8. Graph of evaluation metrics from training using the balanced dataset.

From the table and graph above, there is data on the performance metrics from various dataset combinations with different iterations during the model training. Here are some analyses based on the available data:

- **Differences in Dataset Combinations:** It is evident that each dataset combination exhibits different performances in the measured metrics. For example, if we observe the Precision column, some dataset combinations like "yolov4\_blc\_itr2\_2000" (0.86) and "yolov4\_blc\_itr2\_4000" (0.89) show higher precision values compared to other combinations.
- **Influence of Iterations:** In some cases, increasing the number of iterations consistently improves the measured metrics. For instance, if we consider the dataset combinations "yolov4\_blc\_itr1" and "yolov4\_blc\_itr2," it can be seen that higher numbers of iterations result in better performance in metrics like precision, recall, and F1-score.
- **Interrelation of Metrics:** In certain cases, there is a connection observed between the measured metrics. For example, the dataset combination "yolov4\_blc\_itr2\_4000" exhibits higher values of precision (0.89), recall (0.85), and F1-score (0.87) compared to the dataset combination "yolov4\_blc\_itr2\_1000" (precision: 0.78, recall: 0.62, F1-score: 0.69). This indicates that improvements in precision and recall also contribute to an increase in the F1-score.
- **Dependency on Dataset and Iterations:** The analysis reveals that performance improvement is not solely dependent on the number of iterations but also relies on the dataset combination used. For instance, in the dataset combination "yolov4\_blc\_itr3," increasing the number of iterations does not lead to significant improvements in the measured metrics, particularly in precision and recall.

This analysis shows that performance enhancement cannot be guaranteed solely through increasing the number of iterations. Additionally, other factors such as dataset quality, data variation, and parameter tuning should be considered to improve the overall model performance.

Subsequently, from each cross-validation iteration, the best-performing weight was selected, and the results were

summarized in the cross-validation results table shown in Table XIV and visualized in Figure 9.

TABLE XIV. CROSS-VALIDATION RESULT

Iteration	Precision	Recall	F1-score	mAP @0.5
Iteration 1	0.80	0.78	0.79	75.65%
Iteration 2	0.89	0.85	0.87	92.58%
Iteration 3	0.89	0.77	0.83	77.53%
Iteration 4	0.75	0.57	0.65	61.09%
Iteration 5	0.82	0.68	0.74	72.25%

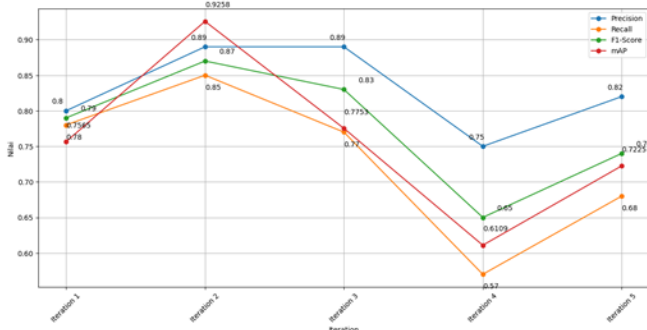


Fig. 9. Graph of evaluation matrix from cross-validation result.

From the cross-validation results above, several analysis can be drawn as follows:

- Precision: Precision measures how accurately the model classifies the positive class. The average precision from five iterations is 0.83, with the highest value reaching 0.89 in the second iteration. This indicates that the model tends to perform well in identifying the positive class.
- Recall: Recall measures how well the model recognizes instances of the positive class. The average recall from five iterations is 0.73, with the highest value reaching 0.85 in the second iteration. Although the average recall is relatively high, it should be noted that there is variation in recall values between different iterations.
- F1-score: The F1-score is a combined measure of precision and recall. The average F1-score from five iterations is 0.76, indicating a balanced performance between precision and recall in classifying the positive class.
- mAP @0.5: The mAP (mean Average Precision) at threshold 0.5 is a commonly used evaluation metric in object detection tasks. The average mAP from five iterations is 76.62%, with the highest value reaching 92.58% in the second iteration. This shows that the model has a good ability to detect objects with confidence levels that meet this threshold.

Thus, based on the cross-validation results, the model demonstrates good performance in classifying the positive class with a relatively high level of accuracy. Consequently, we select the weight associated with the best metric, which is the

metric from iteration 2: "yolov4\_blc\_itr2\_4000," for further analysis and comparison in our research.

### C. Comparison of Imbalanced Dataset and Balanced Dataset Experiment

After conducting training experiments on both imbalanced and balanced datasets, the best-performing weights from each dataset were selected for comparison. The comparison of these weights is presented in Table XV and the graph in Figure 10.

TABLE XV. COMPARISON OF EVALUATION METRICS FOR TRAINING ON IMBALANCED AND BALANCED DATASETS

Name of Weight	Dataset Type	Precision	Recall	F1-score	mAP @0.5
yolov4_imb_395_4000	Imbalanced	0.87	0.83	0.85	0.8767
yolov4_blc_itr2_4000	Balanced	0.89	0.85	0.87	0.9258

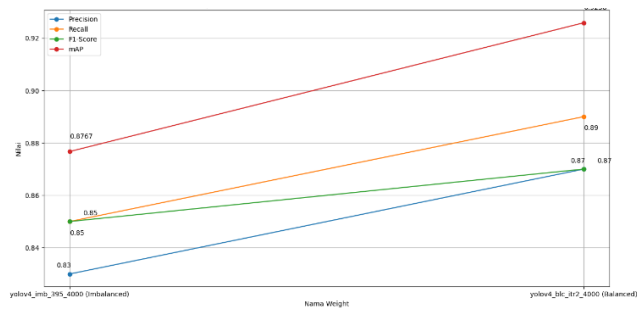


Fig. 10. Comparison graph of evaluation metrics for training on imbalanced and balanced datasets.

- Precision: The model with the balanced dataset has slightly higher precision (0.89) compared to the model with the imbalanced dataset (0.87). Precision measures how accurately the model classifies the positive class, and higher results indicate that the model with the balanced dataset is better at accurately recognizing the positive class.
- Recall: Recall in the model with the balanced dataset (0.85) is slightly lower than the model with the imbalanced dataset (0.83). Recall measures how well the model recognizes instances of the positive class. Although recall in the model with the balanced dataset is higher, the difference is not significant.
- F1-score: The model with the balanced dataset (0.87) has a higher F1-score compared to the model with the imbalanced dataset (0.85). F1-score is a measure of the harmonic mean between precision and recall, and the higher result in the model with the balanced dataset indicates better overall performance in classifying the positive class.
- mAP @0.5: The model with the balanced dataset has a higher mAP @0.5 (0.9258) compared to the model with the imbalanced dataset (0.8767). mAP @0.5 is a commonly used evaluation metric in object detection tasks, and the higher result in the model with the balanced dataset indicates better ability to detect objects with confidence levels that meet the threshold.

Based on the performance evaluation results in the table and graph above, the weight "yolov4\_blc\_itr2\_4000" using the balanced dataset outperforms the weight "yolov4\_imb\_395\_4000" using the imbalanced dataset in terms of precision, recall, F1-score, and mAP @0.5. The model with the balanced dataset has better ability to accurately classify the positive class, with higher F1-score and mAP @0.5.

In addition to better metric results, using the balanced dataset also provides several other advantages. By using a balanced dataset, we can avoid bias in the model as the dataset reflects an even distribution of data. Additionally, a balanced dataset can provide more balanced classification results and better control over prediction errors.

Therefore, based on the comparison results and the advantages obtained, the model with the balanced dataset ("yolov4\_blc\_itr2\_4000") can be chosen as the best in classifying SC-Good-Condition and SC-Bad-Condition.

**D. Model Performance Evaluation and Analysis**

After selecting the "yolov4\_blc\_itr2\_4000" weight using the balanced dataset as the best training result based on the metrics, the performance of the model will be further evaluated and analyzed using the confusion matrix.

From testing 40 validation images on the balanced dataset, a confusion matrix is constructed as a tool to analyze the trained classification model. Figure 11 displays the confusion matrix of the selected weight.

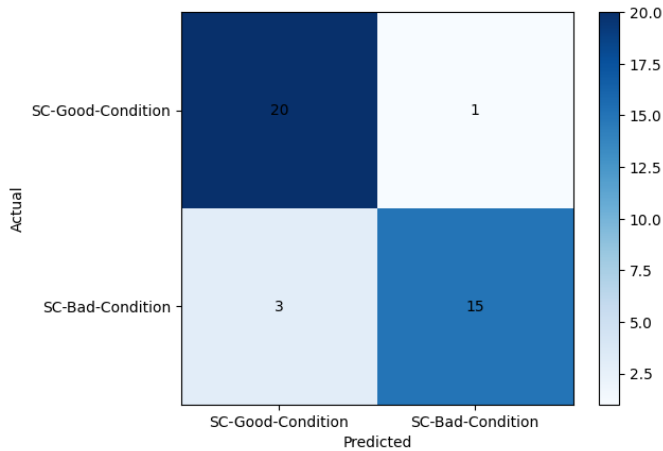


Fig. 11. Confusion matrix of the selected weight.

Based on the confusion matrix above, the model weight can be evaluated as follows:

- TP (True Positives): There are 20 images correctly detected as SC-Good Condition.
- TN (True Negatives): There are 15 images correctly detected as SC-Bad Condition.
- FP (False Positives): There is 1 image wrongly detected as SC-Bad Condition, whereas it should be SC-Good Condition.

- FN (False Negatives): There are 3 images misclassified as SC-Good Condition, whereas they should be SC-Bad Condition.

To improve the model's performance in handling false positives (FP) and false negatives (FN) cases, several additional improvements can be implemented:















- Increase the Amount of Data: Collecting more submarine cable images with a wider variation. Having a larger and more representative dataset allows the model to learn more complex patterns and enhance its detection capabilities for cases of false positives and false negatives.
- Further Data Augmentation: Perform data augmentation on submarine cable images with even broader variations, such as rotation, translation, zooming, cropping, and other distortions. This will help the model recognize various possible conditions in submarine cables and improve its adaptability.
- Hyperparameter Configuration Tuning: Perform hyperparameter tuning on the YOLOv4 model. Some hyperparameters that can be examined include learning rate, max\_batches, batch size, subdivision size, input image size, and other parameters related to the YOLOv4 architecture.



**E. Detection Test on Submarine Cable Image Samples**

In this stage, a detection test is conducted on 20 sample submarine cable images, comprising 10 images with good condition and 10 images with bad condition. These images are outside of the balanced dataset used for training. The results of the detection test on the sample images are shown in Table XVI. Additionally, the Average Precision (AP) graphs for the SC-Good-Condition and SC-Bad-Condition classes are presented in Figure 12 and Figure 13, respectively.

TABLE XVI. RESULTS OF THE DETECTION TEST ON SUBMARINE CABLE IMAGE SAMPLES

File Name	Detection Image Result	Class	AP @0.5
Sample01.png		SC-Good-Condition	0.87
Sample02.png		SC-Good-Condition	0.94
Sample03.png		SC-Good-Condition	0.87
Sample04.png		SC-Good-Condition	0.82

Sample05.png		SC-Good-Condition	0.94
Sample06.png		SC-Good-Condition	0.91
Sample07.png		SC-Good-Condition	0.95
Sample08.png		SC-Good-Condition	0.82
Sample09.png		SC-Good-Condition	0.96
Sample10.png		SC-Good-Condition	0.97
Sample11.png		SC-Bad-Condition	0.82
Sample12.png		SC-Bad-Condition	0.90
Sample13.png		SC-Bad-Condition	0.94
Sample14.png		SC-Bad-Condition	0.83
Sample15.png		SC-Bad-Condition	0.87
Sample16.png		SC-Bad-Condition	0.88
Sample17.png		SC-Bad-Condition	0.66
Sample18.png		SC-Bad-Condition	0.98

Sample19.png		SC-Bad-Condition	0.96
Sample20.png		SC-Bad-Condition	0.91

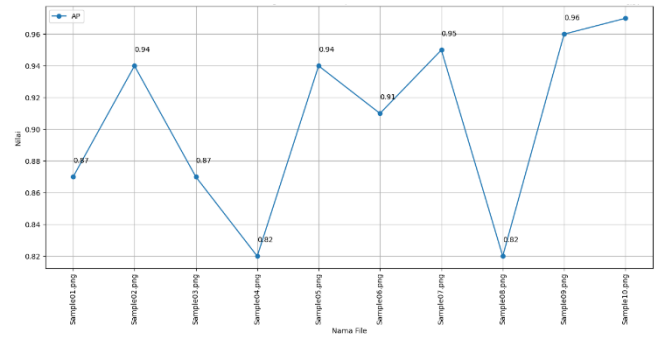


Fig. 12. Graph of average precision results for the detection test on SC-Good-Condition class image samples.

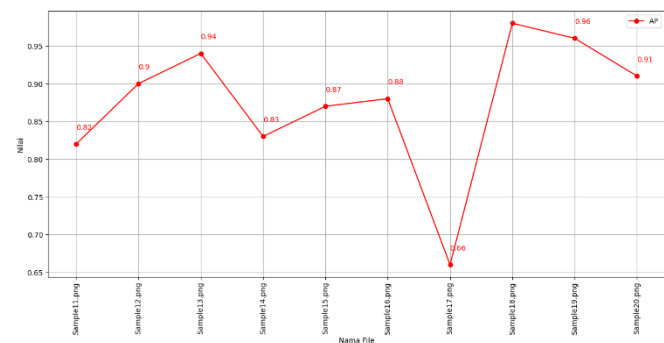


Fig. 13. Graph of Average Precision results for the detection test on SC-Bad-Condition class image samples.

Based on the table and graphs above, the AP (Average Precision) values indicate how well the model can detect and classify objects with high accuracy. The higher the AP value, the better the detection performance on those images. It can be observed that some SC-Good-Condition and SC-Bad-Condition images have high AP values, such as Sample10.png (SC-Good-Condition) with AP 0.97 and Sample18.png (SC-Bad-Condition) with AP 0.98. This indicates that the model has a good ability to detect and classify both conditions of submarine cables. There is variation in the detection performance among the submarine cable images. Some images achieve high AP values, showing accurate detection, while others obtain lower AP values, suggesting possible difficulties in detection for those images. Therefore, it is necessary to evaluate the causes of the low AP values and identify factors that may influence the detection results in those images. This may involve visual analysis and further investigation of the images to discover patterns or difficulties that the model may encounter in accurately classifying submarine cable images.

Here are some assumptions about the factors that may influence the detection results with low AP values on the submarine cable images:

- **Blurry or Unclear Image Quality:** Images with blurry or unclear quality can cause difficulties in detecting submarine cables. The lines or contours of the submarine cables may not be clearly visible, making it challenging for the model to classify accurately.
- **Inappropriate Image Scale or Size:** Image size that is either too small or too large can affect the model's ability to detect submarine cables accurately. Images that are too small may lead to the submarine cable object being too tiny to detect, while images that are too large may cause the details of the submarine cable object to be lost or distorted.
- **Variations in Lighting or Image Contrast:** Differences in lighting or contrast in submarine cable images can affect the model's ability to recognize and classify the submarine cables accurately. Low lighting or low contrast can make it difficult to see the details of the submarine cable object, leading to detection challenges.
- **Presence of Confusing Objects:** The presence of other objects that have visual similarities with submarine cables, such as other cables or structural elements, or the presence of other objects obstructing the submarine cable, such as fish, rocks, etc., can confuse the model and disrupt the process of accurate detection and classification.
- **Variations in Submarine Cable Shapes or Types:** Images in the dataset may have variations in the shape or type of submarine cables, which can pose challenges for the model in recognizing these variations. The model may struggle to understand the variations in texture, shape, or size of different types of submarine cables.

*F. GUI Functionality Test*

In this stage, the functionality of the GUI is tested by conducting non-real-time submarine cable detection on several videos presented in Table IX. The results of the GUI functionality test are presented in Table XVII and graph images in Figure 14 for the testing on the SampleVideo1.mov file.

TABLE XVII. RESULTS OF GUI FUNCTIONALITY TEST ON SAMPLEVIDEO1.MOV

Video Duration	GUI Processing Duration	Detected Class	AP Value @0.5
00:01	00:06	SC-Bad-Condition	0.75
00:01	00:11	SC-Bad-Condition	0.85
00:02	00:16	SC-Bad-Condition	0.89
00:03	00:20	SC-Bad-Condition	0.91
00:04	00:25	SC-Bad-Condition	0.97
00:05	00:29	SC-Bad-Condition	0.97

00:06	00:33	SC-Bad-Condition	0.98
00:06	00:38	SC-Bad-Condition	0.93
00:07	00:42	SC-Bad-Condition	0.95
00:08	00:47	SC-Bad-Condition	0.98
00:09	00:51	SC-Bad-Condition	0.98
00:09	00:56	SC-Bad-Condition	0.98
00:10	01:00	SC-Bad-Condition	0.99
00:11	01:04	SC-Bad-Condition	0.99
00:12	01:09	SC-Bad-Condition	0.97
00:12	01:14	SC-Bad-Condition	0.96
00:14	01:18	SC-Bad-Condition	0.81
00:14	01:22	SC-Bad-Condition	0.73
00:15	01:26	SC-Bad-Condition	0.95
00:16	01:31	SC-Bad-Condition	0.98
00:17	01:35	SC-Bad-Condition	0.91
00:18	01:39	SC-Bad-Condition	0.82
00:18	01:44	SC-Good-Condition	0.93
00:19	01:48	SC-Good-Condition	0.94
00:20	01:52	SC-Good-Condition	0.91
00:21	01:57	SC-Good-Condition	0.94
00:25	02:16	SC-Bad-Condition	0.89
00:26	02:20	SC-Bad-Condition	0.90
00:27	02:25	SC-Bad-Condition	0.50
00:28	02:29	SC-Bad-Condition	0.85
00:30	02:38	SC-Good-Condition	0.75
00:30	02:42	SC-Good-Condition	0.78
00:31	02:47	SC-Good-Condition	0.79
00:32	02:56	SC-Good-Condition	0.66
00:34	03:00	SC-Bad-Condition	0.73
00:34	03:05	SC-Good-Condition	0.54
00:39	03:24	SC-Bad-Condition	0.82
00:39	03:28	SC-Bad-Condition	0.85
00:41	03:33	SC-Bad-Condition	0.32
00:42	03:42	SC-Bad-Condition	0.76
00:44	03:51	SC-Bad-Condition	0.89
00:44	03:55	SC-Bad-Condition	0.85
00:45	04:00	SC-Bad-Condition	0.90
00:47	04:09	SC-Bad-Condition	0.75
00:47	04:13	SC-Bad-Condition	0.71
00:48	04:17	SC-Bad-Condition	0.79
00:48	04:22	SC-Bad-Condition	0.79
00:50	04:26	SC-Good-Condition	0.41
00:51	04:35	SC-Good-Condition	0.78
00:52	04:40	SC-Good-Condition	0.54

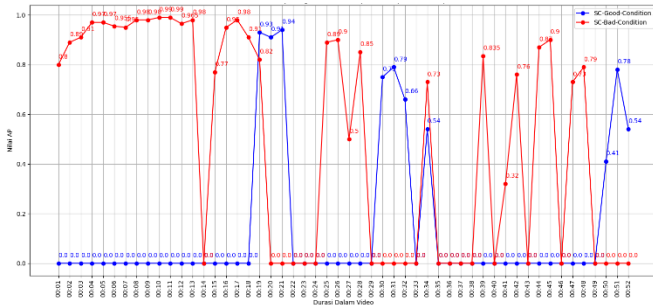


Fig. 14. Graph of GUI functionality test results on SampleVideo1.mov.

Based on the table and graph above, the analysis is as follows:

- The GUI successfully detected SC-Good-Condition at the video timestamps: 00:18, 00:19, 00:20, 00:21, 00:30, 00:31, 00:32, 00:34, 00:50, 00:51, and 00:52. The GUI successfully detected SC-Good-Condition 11 times out of a total of 50 detection frames.
- The GUI successfully detected SC-Bad-Condition at the video timestamps: 00:01, 00:02, 00:03, 00:04, 00:05, 00:06, 00:07, 00:08, 00:09, 00:10, 00:11, 00:12, 00:13, 00:15, 00:16, 00:17, 00:18, 00:19, 00:25, 00:26, 00:27, 00:28, 00:34, 00:39, 00:41, 00:42, 00:44, 00:45, 00:47, and 00:48. The GUI successfully detected SC-Bad-Condition 30 times out of a total of 50 detection frames.
- Several frames are at the same second, such as frames of SC-Bad-Condition at video timestamps: 00:01, 00:06, 00:09, 00:12, 00:14, 00:18, 00:34, 00:39, 00:44, 00:47, and 00:48, which were detected twice at each second. As for SC-Good-Condition frames, there is only one video timestamp (00:30) with two frames detected in that second.
- Several frames are at the same second but have different detection classes. At video timestamps: 00:18 and 00:34, each has 2 frames with different detection classes, one frame of SC-Good-Condition, and one frame of SC-Bad-Condition.
- To calculate the GUI's FPS (Frames per Second) performance for SampleVideo.mov, the formula in equation (V) is used:

$$FPS = \frac{50 \text{ frame}}{280 \text{ detik}} = 0.178 \text{ fps}$$

From the calculations above, the GUI performance for detecting SampleVideo1.mov still has a low FPS, taking a total time of 4 minutes and 40 seconds (280 seconds).

- To calculate the mAP (mean Average Precision) performance, the formula in equation (IV) is used. The results of mAP calculations for each class in the GUI functionality test on VideoSample01.mov are shown in Table XVIII.

TABLE XVIII. RESULTS OF MAP CALCULATION FOR GUI FUNCTIONALITY TEST ON VIDEOSAMPLE01.MOV

Class Name	mAP @0.5
SC-Good-Condition	0.725
SC-Bad-Condition	0.861

Based on the table above, the detection and classification results of submarine cables using the GUI have achieved a good mAP value.

Next, the functionality test continues for the video SampleVideo2.mov. The results of the GUI functionality test for SampleVideo2.mov are presented in Table XIX and graph images in Figure 15.

TABLE XIX. RESULTS OF GUI FUNCTIONALITY TEST ON SAMPLEVIDEO2.MOV

Video Duration	GUI Processing Duration	Detected Class	AP Value @0.5
00:02	00:07	SC-Good-Condition	0.48
00:02	00:12	SC-Good-Condition	0.95
00:02	00:16	SC-Good-Condition	0.56
00:03	00:20	SC-Good-Condition	0.79
00:04	00:24	SC-Good-Condition	0.35
00:11	00:53	SC-Good-Condition	0.45
00:11	00:58	SC-Good-Condition	0.43
00:11	01:02	SC-Good-Condition	0.92
00:12	01:06	SC-Good-Condition	0.86
00:13	01:11	SC-Good-Condition	0.51
00:14	01:15	SC-Good-Condition	0.49
00:14	01:19	SC-Good-Condition	0.29
00:20	01:35	SC-Good-Condition	0.27
00:30	02:31	SC-Good-Condition	0.56
00:38	03:08	SC-Bad-Condition	0.39
00:44	03:43	SC-Bad-Condition	0.58
00:50	04:09	SC-Bad-Condition	0.41
00:51	04:14	SC-Bad-Condition	0.72
00:51	04:18	SC-Bad-Condition	0.94
00:52	04:22	SC-Bad-Condition	0.98
00:53	04:27	SC-Bad-Condition	0.90
00:54	04:31	SC-Bad-Condition	0.94
00:54	04:36	SC-Bad-Condition	0.74

Based on the table and graph above, the analysis is as follows:

- The GUI successfully detected SC-Good-Condition at the video timestamps: 00:02, 00:03, 00:04, 00:11, 00:12, 00:13, 00:14, 00:20, and 00:30. The GUI successfully detected SC-Good-Condition 9 times out of a total of 23 detection frames.



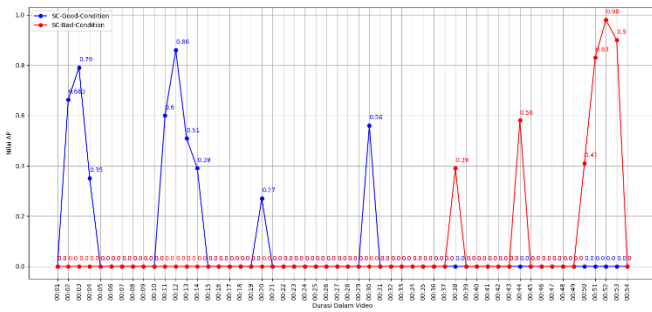


Fig. 15. Graph of GUI functionality test results on SampleVideo2.mov.

- The GUI successfully detected SC-Bad-Condition at the video timestamps: 00:38, 00:44, 00:50, 00:51, 00:52, and 00:54. The GUI successfully detected SC-Bad-Condition 6 times out of a total of 23 detection frames.
- Several frames are at the same second, such as SC-Bad-Condition frames at video timestamps: 00:51 and 00:54, which were detected twice at each second. SC-Good-Condition frames at video timestamps: 00:13 and 00:14 were detected three times in each second.
- To calculate the GUI's FPS (Frames per Second) performance for SampleVideo.mov, the formula in equation (V) is used:

$$FPS = \frac{23 \text{ frame}}{276 \text{ detik}} = 0.083 \text{ fps}$$

Based on the calculations above, the GUI performance for detecting SampleVideo2.mov still has a low FPS, taking a total time of 4 minutes and 36 seconds (276 seconds).

- From the testing of SampleVideo1.mov and SampleVideo2.mov, both GUI FPS performances are relatively low. The following are some assumptions that may cause the low FPS performance:
  - Model Complexity: The low FPS may be caused by the complexity of the YOLOv4 detection model used. Models with many layers and parameters may require longer processing time.
  - Computational Load: Object detection using the YOLOv4 model requires intensive processing and consumes a lot of computational power. This can reduce processing speed and result in low FPS.
  - Hardware Limitations: The use of less powerful hardware, such as a CPU with low computational power, can affect the GUI's performance and cause low FPS.
- To improve FPS and enhance the GUI's responsiveness in submarine cable detection, several efforts can be made:
  - Model Optimization: Optimize the YOLOv4 detection model by reducing the number of unnecessary layers or parameters and using a lighter model.

- Use More Powerful Hardware: Use GPUs with parallel processing capabilities to improve processing speed and FPS.
- Resolution Reduction: Reduce the video resolution or convert the video format to a lighter format to improve FPS.
- Data Streaming: Use data streaming techniques to process videos in real-time and enhance GUI responsiveness.

By implementing the above efforts, it is expected that FPS on the GUI can be increased, making the application more responsive and providing users with a better experience in submarine cable detection. Continuous and iterative evaluation is necessary to ensure optimal performance improvements.

- To calculate the mAP (mean Average Precision) performance, the formula in equation (IV) is used. The results of mAP calculations for each class in the GUI functionality test on VideoSample02.mov are shown in Table XX.

TABLE XX. RESULTS OF MAP CALCULATION FOR GUI FUNCTIONALITY TEST ON VIDEOSAMPLE02.MOV

Class Name	mAP @0.5
SC-Good-Condition	0.554
SC-Bad-Condition	0.681

### G. Comparison of Functionality Test Performance

From the two functionality tests with two different videos, VideoSample01.mov and VideoSample02.mov, two performances were obtained for comparison. The comparison of functionality test performances is presented in Table XXI.

TABLE XXI. COMPARISON OF GUI FUNCTIONALITY TEST RESULTS FOR VIDEOSAMPLE01.MOV AND VIDEOSAMPLE02.MOV

File Name	Video Duration	GUI Processing Time	Number of SC Frame Detected	FPS	mAP @0.5
VideoSample01.mov	00:54	04:40	50 frames	0.178	0.829
VideoSample02.mov	01:00	04:36	23 frames	0.083	0.605

Overall, the GUI has performed well in detecting submarine cables in both videos. This can be seen from the relatively high mean Average Precision (mAP) @0.5 values, which are 0.829 for VideoSample01.mov and 0.605 for VideoSample02.mov. mAP @0.5 measures the object detection accuracy at an Intersection over Union (IoU) threshold of 0.5, and the higher the mAP value, the more accurate the detection results.

However, there are differences in the number of detected submarine cable frames between the two videos. In VideoSample01.mov, the GUI successfully detected 50 SC frames, while in VideoSample02.mov, the number of detected submarine cable frames was only 23. This difference is due to the varying video conditions between the two samples.

VideoSample02.mov has a more blurry, shaky, and often unfocused video quality, making submarine cable detection more challenging. This affects the GUI's performance in detecting SC in that video. Another factor that can affect the number of detected submarine cable frames is the complexity of the background and the presence of other objects that are similar to the submarine cable, causing some submarine cable frames to go undetected.

Regarding FPS, both VideoSample01.mov and VideoSample02.mov have low FPS values. This indicates that the GUI's processing is still relatively slow, resulting in slow detection. Increasing the FPS is one of the efforts that need to be made to improve the quality and efficiency of the GUI in detecting submarine cables.

#### IV. CONCLUSION

In this research, a YOLOv4-based underwater detection system integrated with a Graphical User Interface (GUI) for Remotely Operated Vehicle (ROV) in submarine cable detection has been successfully designed and implemented. Based on the experimental results and analysis, the following are the conclusions drawn from this research:

1) *Development of Submarine Cable Detection Model:* The performance evaluation of the model on the balanced dataset weight showed satisfactory results with precision of 0.89, recall of 0.85, F1-score of 0.87, and mAP of 92.58%. This indicates the model's ability to recognize both classes effectively.

2) *Implementation of Graphical User Interface (GUI):* The performance evaluation of the designed GUI showed promising results. In VideoSample01.mov, the GUI successfully detected 50 frames of submarine cable images with an mAP of 0.829 and GUI FPS of 0.0178. In VideoSample02.mov, the GUI detected 23 frames of submarine cable images with an mAP of 0.605 and GUI FPS of 0.083. The implementation of the GUI with the submarine cable detection model on the ROV successfully reduces dependency on human observation. With this automated system, issues of fatigue and subjective interpretation in identifying submarine cable conditions can be addressed. This provides the benefit of facilitating the submarine cable maintenance process.

In light of the successful development and implementation of the YOLOv4-based underwater detection system integrated with a GUI for ROV in submarine cable detection, there are several key areas for future research and improvements:

1) *Future work* should aim to enhance the scalability and adaptability of the system. This could involve expanding the dataset to encompass a broader range of underwater environments and conditions. Additionally, exploring the integration of machine learning techniques for automatic parameter tuning, especially in varying lighting and water clarity conditions, would further bolster the system's performance and reliability. Furthermore, the system could benefit from the incorporation of real-time anomaly detection

algorithms to promptly identify potential cable issues and facilitate proactive maintenance.

2) *There* is potential to extend the application of this technology to broader marine infrastructure management. Researchers can explore its utility in tasks beyond submarine cable detection, such as pipeline monitoring, marine biodiversity assessment, and archaeological exploration. Adapting the system for these diverse applications could significantly contribute to the advancement of marine sciences and industries. Additionally, research efforts should be directed toward refining the user interface and operator interaction aspects of the GUI to ensure user-friendliness and efficiency. This includes incorporating features that enable operators to annotate and validate detected cable segments, fostering human-machine collaboration for more accurate results. By addressing these areas in future research endeavors, we can further enhance the capabilities and impact of this innovative underwater detection system.

#### REFERENCES

- [1] M. Jamin and A. Sugiyono, "Pengembangan Kelistrikan Nasional."
- [2] Susianti, E., Syahputra, N. A., Wibowo, A. U., & Maria, P. S. (2021). Rancang Bangun Robot Observasi Bawah Air-ROV (Remotely Operated Vehicle) Menggunakan Arduino UNO. Jurnal Elektro dan Mesin Terapan, 7(2), 136-146.
- [3] Saputro, B. S., Djunarsjah, E., Setiyadi, J., & Negara, A. K. (2015). Pengoperasian Remotely Operated Vehicle (ROV) Mendukung Pekerjaan Bawah Air (Studi Kasus Pendeteksian Kabel Bawah Laut Menggunakan ROV H800 Di Perairan Selat Bangka Belitung): Remotely Operated Vehicle (ROV) Operation Supports Underwater Work (Case Study of Detecting Submarine cables Using ROV H800 in the Waters of the Bangka Belitung Strait). Jurnal Hidropilar, 1(2), 95-111.
- [4] Noyes, R. Y. (1994). Inspection methods for underwater cables (Doctoral dissertation, National Technical Information Service).
- [5] Zhang, M., Xu, S., Song, W., He, Q., & Wei, Q. (2021). Lightweight underwater object detection based on yolo v4 and multi-scale attentional feature fusion. Remote Sensing, 13(22), 4706.
- [6] Rosli, M. S. A. B., Isa, I. S., Maruzuki, M. I. F., Sulaiman, S. N., & Ahmad, I. (2021, August). Underwater animal detection using YOLOV4. In 2021 11th IEEE International Conference on Control System, Computing and Engineering (ICCSCE) (pp. 158-163). IEEE.
- [7] Zhang, C., Zhang, G., Li, H., Liu, H., Tan, J., & Xue, X. (2023). Underwater target detection algorithm based on improved YOLOv4 with SemiDSCov and FIoU loss function. Frontiers in Marine Science, 10, 1153416.
- [8] García-Valdovinos, L. G., Salgado-Jiménez, T., Bandala-Sánchez, M., Nava-Balanzar, L., Hernández-Alvarado, R., & Cruz-Ledesma, J. A. (2014). Modelling, design and robust control of a remotely operated underwater vehicle. International Journal of Advanced Robotic Systems, 11(1), 1.
- [9] Li, Y., Zhang, X., & Shen, Z. (2022). YOLO-Submarine cable: An Improved YOLO-V3 Network for Object Detection on Submarine cable Images. Journal of Marine Science and Engineering, 10(8), 1143.
- [10] Matsumoto, S., & Ito, Y. (1995, October). Real-time vision-based tracking of submarine-cables for AUV/ROV. In 'Challenges of Our Changing Global Environment'. Conference Proceedings. OCEANS'95 MTS/IEEE (Vol. 3, pp. 1997-2002). IEEE.
- [11] Fatan, M., Daliri, M. R., & Shahri, A. M. (2016). Underwater cable detection in the images using edge classification based on texture information. Measurement, 91, 309-317.
- [12] Burnett, D. R., Beckman, R., & Davenport, T. M. (Eds.). (2013). Submarine cables: the handbook of Law and Policy. Martinus Nijhoff Publishers.

- [13] Adesokan, A. A. (2021). Covid-19 Control: Face Mask Detection Using Deep Learning for Balanced and Unbalanced Dataset. Available at SSRN 4181373.
- [14] Confusion Matrix - an overview | ScienceDirect Topics. (n.d.). Confusion Matrix - an Overview | ScienceDirect Topics. <https://doi.org/10.1016/B978-0-12-818366-3.00005-8>
- [15] Sharma, D. K., Chatterjee, M., Kaur, G., & Vavilala, S. (2022). Deep Learning applications for disease diagnosis. In Deep Learning for medical applications with unique data (pp. 31-51). Academic Press.

# Research on Clothing Color Classification Method based on Improved FCM Clustering Algorithm

Jinliang Liu

School of Creative Design, Shanghai Industry of Commerce & Foreign Languages, Shanghai 201399, China

**Abstract**—In the apparel industry, apparel color is an important factor to enhance the market competitiveness of enterprise products. However, the current prediction samples of clothing fashion color styling information do not incorporate practical cutting-edge fashion information. Therefore, Self-adaptive Weighted Kernel Function (SWK) has been introduced to traditional Fuzzy C-Means (FCM) clustering algorithms. After improvement, the SWK-FCM clustering algorithm is obtained, which enhances the classification ability of fashion colors and hue. Two prediction models have been developed using the finalized data of the International Fashion Color Committee, along with the SWK-FCM clustering algorithm. The models have been tested via experiments to verify their accuracy. The experimental results show that the classification coefficients of SWK-FCM clustering algorithm are 0.9553 and 0.9258 under 5% Gaussian noise. They are higher than those of FCM (0.7063) and FLICM (0.8598). The classification entropy is lower than that of the comparison algorithm, while the same results are presented under other conditions and in the actual experiments. In addition, the overall MSE of the GM (1, 1) prediction model using the final case information is 0.00028, which is close to the order of  $10^{-4}$ . The MSE value of the BP neural network prediction model using the final case information ranges from 0.000529 to 0.011025. Overall, the clustering algorithm of SWK-FCM has good classification performance. Additionally, the GM (1,1) model based on SWK-FCM has better prediction results, which can be effectively applied in practical clothing color classification and popular color prediction.

**Keywords**—Fuzzy clustering; SWK-FCM; fashion color scheme; Gaussian noise

## I. INTRODUCTION

The development of the economy has driven changes in the overall consumption concept of the people. People's daily needs for clothing are gradually shifting towards pursuing quality and fashion [1]. Clothing color is an important element in improving the competitiveness of goods. Popular colors are the focus of attention in the clothing industry, and their effective prediction can promote the effective development of clothing enterprises [2]. Zhu D et al. constructed two neural models of perceptual data based on deep learning for the problem of popular color trends of Japanese women's clothing kawaii style [3]. In response to the difficulty of fashion designers in achieving realistic restoration of Yi clothing color images, Zhu Hai et al. proposed a Yi clothing color extraction classification and trend prediction based on the K-means algorithm [4]. Garcia C C C proposed a consumer preference data based on material culture for fashion culture prediction in response to the trend of fashion clothing color trends [5]. The

current methods for predicting fashion color trends are still in the exploratory stage, resulting in low prediction accuracy, high time consumption, and high actual cost. As a clustering algorithm, FCM algorithm can effectively solve the classification time and improve the accuracy of prediction. However, traditional FCM does not perform well in actual fashion color classification prediction. At the same time, although the color quantization space method among many current methods can predict clothing color systems based on final information, it has high requirements for data extraction and low practicality. The grey model in the popular color prediction model has low requirements for dyeing sample data, but the actual prediction accuracy is not high enough. In this context, this study improves the FCM clustering algorithm and obtains an FCM clustering algorithm based on mean space constraints and adaptive weighted kernel functions (Self-adaptive Weighted Kernel Fuzzy C-Means, SWK-FCM). Based on this, the Grey Model (1,1), GM (1,1) and Back Propagation (BP) neural network prediction model are constructed. The purpose of the model is to improve the classification quality of apparel hues and achieve effective prediction of apparel fashion colors, so as to give relevant enterprises the corresponding guidance. In addition, the application of data prediction in the textile industry is still in an exploratory state, and the application of classification methods in the clothing industry is relatively shallow. Therefore, the study of using SWK-FCM clustering algorithm to classify clothing hue is innovative.

This paper is divided into five sections in total. The Section I is a summary and discussion of the current research on clothing color classification at home and abroad. Section II is the related work. Section III analyzes the clothing color classification method proposed on the basis of the improved FCM clustering algorithm. Section IV is to analyze the performance of this method. Section V concludes the paper.

## II. RELATED WORK

Fashion colors are the most important part of fashion elements, and they are pivotal in enhancing brand image and sales [6]. Therefore, achieving effective prediction of apparel fashion colors can bring higher business value to stakeholders and also promote the development of the industry. However, the prerequisite for realizing fashion color prediction for apparel is the effective classification of apparel hues [7]. Based on this, a wide range of domestic and foreign scholars have conducted in-depth research on it. Zhou Ze et al. proposed a new clothing classification method based on parallel convolutional neural networks for color tone recognition and classification. This method improved the

stability and accuracy of clothing classification on the basis of effectively extracting clothing tone features in [8]. Mushtaque S et al. constructed an effective classification platform for online shopping of women's clothing in Pakistan based on the collection of top women's clothing brands in Pakistan [9]. Wu D et al. proposed a new method for apparel attribute recognition and prediction, and used it to classify apparel hues, thus improving the recognition efficiency and classification accuracy based on the constructed relational model. The accuracy of classification was improved based on the relationship model [10]. Dai Y et al. proposed five different experimental schemes to improve the accuracy of fashion color database and fashion color classification based on the collection of relevant data [11]. Han A et al. used machine learning to analyze fashion color data and determine whether fashion designers have used fashion colors proposed by relevant institutions to guide seasonal trends. This method of analyzing and classifying fashion color data effectively enhances the potential of fashion color trend analysis to guide production [12].

In addition, Mau T N et al. effectively created an optimization method for initial clustering based on location-sensitive hashing for fuzzy clustering of categorical data, thus improving the clustering effect while reducing the dimensionality of categorical data and providing help for data trend prediction [13]. Mersch B and Buchel S et al. proposed a pricing heuristic based on machine learning for large-scale optimization of images, which improves the image coloring problem while increasing the prediction accuracy [14]. The issue of sustainable development was comprehensively analyzed through a multi-level perspective system to provide a predictive direction for the strategic transformation of the fashion industry [15-16]. Wang Wei et al. constructed a hybrid color trend prediction model based on genetic algorithm and extreme learning machine to address the issue of future color trend prediction, effectively improving prediction accuracy [17].

From the research of above scholars, the application of prediction of data in textile industry is still in the state of exploration, while the application of classification method in apparel industry is still shallow. Therefore, the research on the classification of apparel color shades using the SWK-FCM clustering algorithm possesses a certain degree of innovation, while the prediction model constructed on its basis can provide objective and scientific support to the stakeholders. In addition, it can also provide guidance for guiding consumers to form correct consumption behaviors while improving the accuracy of apparel trend prediction.

### III. ANALYSIS OF CLOTHING COLOR CLASSIFICATION METHOD BASED ON IMPROVED FCM CLUSTERING ALGORITHM

#### A. Study on Clothing Color Classification based on SWK-FCM Clustering Algorithm

There is a problem of inconsistency between current fashion color case information and actual cutting-edge fashion information. In response to this, this study combined case information and improved the Fuzzy C-Means (FCM) clustering algorithm to construct a fashion color prediction model. Traditional clustering usually refers to the process of

clustering things using similarity. Cluster analysis uses mathematical methods for clustering research, using the degree of correlation between certain features of the tested object as the basis for classification, without considering other prior knowledge. Fuzzy cluster analysis is a clustering analysis method constructed on the basis of fuzzy theory. Its correlation analysis of data is performed on the basis of considering regional characteristics and classifying them [18]. In pattern recognition and image segmentation, this method can be used to describe fuzzy objects more objectively.

FCM clustering algorithm belongs to a kind of fuzzy clustering algorithm, which achieves automatic classification of sample data by optimizing each sample point and determining the class of each sample point [19]. The classification accuracy of the traditional FCM clustering algorithm is very effective in the actual clothing color classification. Therefore, the study obtained the SWK-FCM clustering algorithm by embedding the kernel function in the FCM clustering algorithm and weighting it. Compared with traditional FCM, the SWK-FCM clustering algorithm enhances inter sample features by introducing kernel functions. It effectively prevents misclassification of important details through adaptive weights, and improves the system's noise resistance by introducing spatial functions. Among them, the expression of the objective function of SWK-FCM is shown in Eq. (1).

$$T = 2 \sum_{i=1}^a \sum_{j=1}^n b_i^m \lambda_{ij}^m (1 - K(x_j, u_i)) + 2\gamma \sum_{i=1}^a \sum_{j=1}^n b_i^m \lambda_{ij}^m (1 - K(\bar{x}_j, u_i)) \quad (1)$$

In equation (1),

$$T = 2 \sum_{i=1}^a \sum_{j=1}^n b_i^m \lambda_{ij}^m (1 - K(x_j, u_i)) + 2\gamma \sum_{i=1}^a \sum_{j=1}^n b_i^m \lambda_{ij}^m (1 - K(\bar{x}_j, u_i))$$

denotes the objective function.  $i$  denotes the number of clusters, whose maximum value is  $a$ .  $j$  denotes the number of elements in the sample, whose maximum value is  $n$ .  $b_i$  denotes the dynamic weight of a class, which can express the importance of a certain class.  $\lambda$  denotes the affiliation degree.  $m$  denotes the fuzzy factor.  $J(x, u)$  denotes the inner product kernel function.  $\gamma$  denotes the control parameter. On this basis, the computational expressions of the clustering center and the affiliation function can be obtained by minimizing the objective function by the Lagrange multiplier method under the corresponding constraints. Among them, the expressions of the constraints are shown in Eq. (2).

$$\begin{cases} \sum_{i=1}^a b_i = 1 \\ \sum_{i=1}^a \lambda_{ij} = 1 \end{cases} \quad (2)$$

Therefore, the expression of the cluster center  $u_i$  is calculated as shown in Eq. (3).

$$u_i = \frac{\sum_{j=1}^n \lambda_{ij}^m J(x_j, u_i)(x_j, \bar{u}_j)}{(1 + \gamma) \sum_{j=1}^n \lambda_{ij}^m J(x_j, u_i)} \quad (3)$$

Similarly, the expression for the calculation of dynamic weights is shown in Eq. (4).

$$b_i = \frac{\sum_{j=1}^n \lambda_{ij}^m (1 - J(x_j, u_i))^{\frac{1}{m-1}}}{\sum_{i=1}^a \left( \sum_{j=1}^n \lambda_{ij}^m (1 - J(x_j, u_i))^{\frac{1}{m-1}} \right)^{\frac{1}{m-1}}} \quad (4)$$

Finally, the expression of the affiliation function is calculated as shown in Eq. (5).

$$\lambda_{ij} = \frac{b_i \left( 1 - J(x_j, u_i) + \gamma \left( 1 - J(x_j, \bar{u}_i) \right) \right)^{\frac{1}{m-1}}}{\sum_{k=1}^a b_k \left( 1 - J(x_k, u_i) + \gamma \left( 1 - J(x_k, \bar{u}_i) \right) \right)^{\frac{1}{m-1}}} \quad (5)$$

Therefore, the specific steps of the SWK-FCM clustering algorithm constructed in the study are shown in Fig. 1.

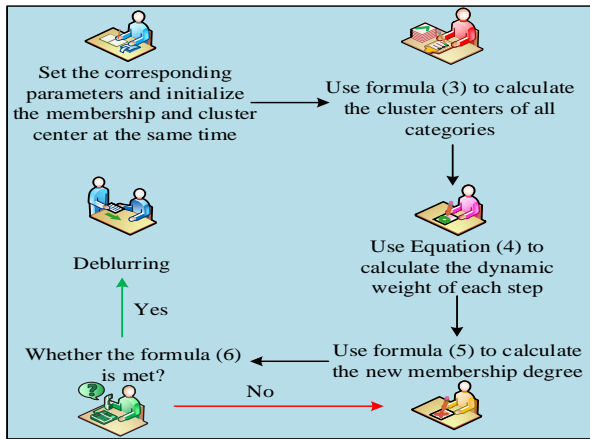


Fig. 1. Detailed steps of SWK-FCM clustering algorithm.

In Fig. 1, the SWK-FCM algorithm firstly sets the corresponding parameters and initializes the affiliation and clustering centers at the same time. Secondly, it calculates the clustering centers of all categories and the dynamic weights of each step using Eq. (3) and Eq. (4). Then it updates the affiliation and calculates the new affiliation using Eq. (5). It continuously updates the affiliation and calculates the new affiliation until it satisfies Eq. (6). The final step is defuzzification, thus completing the classification. Among them, the determination constraints embodied in the steps are shown in Eq. (6).

$$\max \left\{ \left| \lambda_{ij}^{old} - \lambda_{ij}^{new} \right| \right\} < \delta \quad (6)$$

In Eq. (6),  $\lambda_{ij}^{old}$  denotes the old affiliation.  $\lambda_{ij}^{new}$  denotes the new affiliation.  $\delta$  denotes the threshold value. In

addition, in the final deblurring equation, the classification of each pixel point needs to be achieved based on the corresponding point equation, which is expressed as shown in Eq. (7).

$$v = \arg \max x_i \{ \lambda_{ij}, i = 1, 2, \dots, a \} \quad (7)$$

In Eq. (7),  $v$  indicates the category to which it belongs. Since the color system of different popular colors itself is different, this leads to different methods of color classification. Therefore, in the actual process of popular color data processing, it is necessary to determine the popular color hue classification method. Based on Eq. (1) to (7), the proposed SWK-FCM algorithm can automatically classify the color hues of clothing fashions and gradually determine the relevant intervals of color hue values for further processing of the data. Its process in apparel color clustering is similar to Fig. 1. However, there is a slight difference in the last step. That is, according to Eq. (1), the objective function value is calculated and the minimum error value is set. Until the objective function value is less than the given minimum error value or the number of iterations exceeds the limit, the algorithm process ends. The expression of the minimum error value calculation is shown in Eq. (8).

$$\varepsilon = \frac{|T_t - T_{t-1}|}{T_t} \quad (8)$$

In Eq. (8),  $\varepsilon$  denotes the minimum error value.  $t$  denotes the number of iterations. By using the SWK-FCM clustering algorithm to cluster the hue and return to the hue center and hue affiliation matrix, the hue value interval of each color class of the garment can be obtained after the corresponding processing of the affiliation matrix. The individual unclustered hue values are automatically assigned with the hue value interval that spans the adjacent interval. At the same time, the frequency of each type of hue in different years can be obtained statistically. The corresponding formula is used to calculate the proportion. The calculation expression is shown in Eq. (9).

$$P_{i'} = \frac{f_{i'}}{F} \times 100\% \quad (9)$$

In Eq. (9),  $P_{i'}$  indicates the proportion of colors in the  $i'$  category in different years.  $f_{i'}$  indicates the frequency of colors in the  $i'$  category.  $F$  indicates the total number of colors announced in the fashion color scheme of a certain year.

### B. Construction of a Color Prediction Model for Fashionable Clothing Colors based on Final Case Information

After elaborating the SWK-FCM clustering algorithm, the processed data can be brought into GM (1,1) with BP neural network to achieve the prediction of clothing process color shades. Before that, it is necessary to design the process of garment process color trend prediction using the finalized information. The study uses the information of the apparel fashion color case issued by the International Color Council as the actual data source, and uses the SWK-FCM clustering

algorithm to classify the color shades of apparel, so as to redefine the color intervals. GM (1,1) and BP neural network are used to perform color shade prediction, and the specific process is shown in Fig. 2.

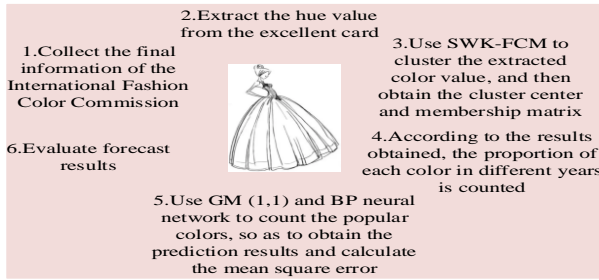


Fig. 2. Fashion color trend prediction process based on final information.

From Fig. 2, the process of predicting the trend of fashion colors using the case information is firstly to collect the historical case information of the International Color Council and extract the color phase values from the color cards. Secondly, it is to cluster the proposed color phase values by using SWK-FCM to obtain the cluster center and affiliation matrix. The next step is to count the proportion of each color in different years based on the results obtained. Then, GM(1,1) and BP neural network are taken to count the popular colors to obtain the prediction results and calculate the mean square error. Finally, evaluating the prediction results. Among them, the main steps of the prediction of the fashion color prediction model using GM (1,1) based on the final case information are shown in Fig. 3.

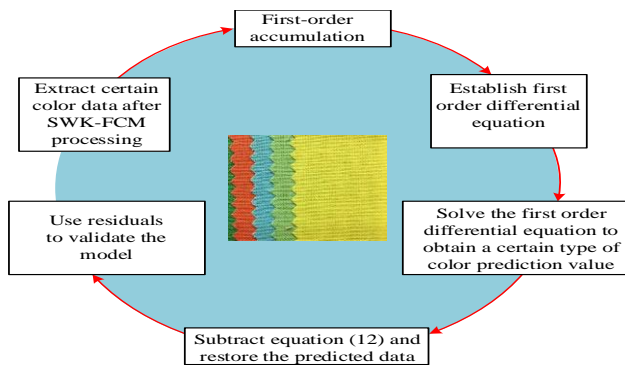


Fig. 3. Main steps of fashion color prediction based on GM (1,1).

From Fig. 3, the main steps are to first extract the original sequence of a certain type of color data after SWK-FCM processing and perform first-order accumulation to generate a first-order sequence; second, to establish a first-order differential equation and solve the first-order differential equation to obtain the predicted value of a certain type of color. Next step is to perform continuous accumulation of the equation to reduce the predicted data. Finally, the residual is used to verify the model. The expression of the original sequence and the first-order cumulative equation is shown in Eq. (10) [20-22].

$$\begin{cases} y^{(0)}(v') = y^{(0)}(1), y^{(0)}(2), \dots, y^{(0)}(v') \\ y^{(1)}(v') = \sum_{i'=1}^{v'} y^{(0)}(i') \end{cases} \quad (10)$$

In Eq. (10),  $y^{(0)}(v)$  denotes the original sequence.  $y^{(1)}(v)$  denotes the first-order sequence.  $i''$  denotes the color category whose maximum value is  $v'$ . And the computational expression of the established first-order differential equation is shown in Eq. (11).

$$\frac{dy^{(1)}(v')}{dv} + cy^{(1)}(v') = h \quad (11)$$

Eq. (11),  $c$  denotes the whitening coefficient, which reflects the trend of the variable.  $h$  denotes the endogenous control gray scale. In addition, the expression of the predicted value of a certain type of color obtained by solving this first-order differential equation is shown in Eq. (12).

$$\hat{y}^{(0)}(v'+1) = \left( y^{(0)}(1) - \frac{h}{c} \right) e^{-cv'} + \frac{h}{c} \quad (12)$$

In Eq. (12),  $e$  denotes the natural constant. Therefore, based on the stepwise representation of the GM (1,1) model, Eq. (12) is cumulated. The predicted data expression obtained by this reduction is shown in Eq.(13).

$$\hat{y}^{(0)}(v'+1) = \hat{y}^{(1)}(v'+1) - \hat{y}^{(1)}(v') \quad (13)$$

Finally, the model is validated using the nibbling formula, and the expression of the residual formula is calculated as shown in Eq. (14).

$$\eta(v') = \frac{y^{(0)}(v') - \hat{y}^{(0)}(v')}{y^{(0)}(v')} \quad (14)$$

In Eq. (14),  $\eta(v')$  denotes the relative error. The model is considered to meet the requirement when the relative error meets the corresponding condition, which is expressed in Eq. (15).

$$\begin{cases} |\eta(v')| < 0.1 \\ |\eta(v')| < 0.2 \end{cases} \quad (15)$$

In Eq. (15), when the relative error satisfies the first row of the equation, it means that the GM model achieves good accuracy. When the relative error satisfies the second row of the equation, it means that the GM model achieves general requirements. In addition to the GM model, the apparel popular color hue prediction model constructed by using the final case information also contains the BP neural network prediction model. BP neural network is a one-way propagation multilayer feedforward network, which is trained repeatedly by the samples of the network. Therefore, the weights and thresholds of the network can be gradually reduced on the negative gradient and achieve the expected results to a certain extent [23]. Because there is little color information of popular colors, and after analyzing the prediction results of current popular colors according to different network structures, the study chose a BP neural network for prediction. Its network topology is schematically shown in Fig. 4.

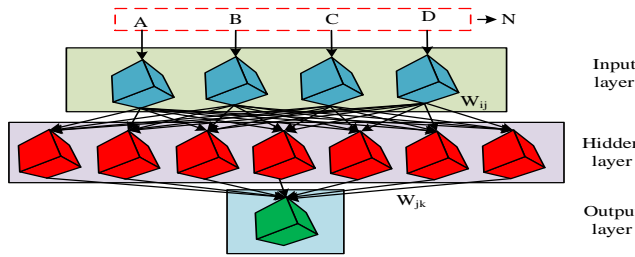


Fig. 4. Topological structure diagram of fashion colors based on BP neural network.

In Fig. 4,  $N$  denotes the color data related to a certain type of clothing color for the previous  $N$  years.  $O$  denotes the target value.  $W_{ij}$  denotes the weight value of the  $i$ -th neuron in the input layer relative to the  $j$ -th neuron in the hidden layer.  $W_{jk}$  denotes the weight value of the  $j$ -th neuron in the hidden layer relative to the  $k$ -th neuron in the output layer. From the figure, the topology of clothing fashion color using BP neural network contains the same three-layer structure. However, in the process of predictive modeling, the data related to each type of color needs to be trained separately. Firstly, the number of nodes in the input layer, output layer and should that layer needs to be set. The study sets the number of nodes in the input layer to 4, which are represented by A, B, C and D respectively. The number of nodes in the implicit layer is 7, and the number of nodes in the output layer is 1. Next, parameters such as excitation function, minimum learning rate, and maximum number of iterations are set. When the error associated with the training exceeds the expected value, it enters the backward propagation of the error, and then the gradient decreasing method is used to gradually correct the hidden layer weights to obtain the expected result. Finally, the completed training model is used to predict the color correlation data for the next year.

#### IV. PERFORMANCE ANALYSIS OF THE PREDICTION MODEL CONSTRUCTED BASED ON SWK-FCM CLUSTERING ALGORITHM

To verify the effectiveness of the garment fashion color shade prediction model constructed in the study, the study analyzed it using experiments. This study first evaluated the quality of clothing color classification using the SWK-FCM

clustering algorithm. Then, the fuzzy local information c-means (FLICM) clustering algorithm was introduced and compared with SWK-FCM clustering algorithm and traditional FCM clustering algorithm [24]. Before the experiment, the threshold value is set to 0.00001, the fuzzy factor is 2, the control parameter is 3.8, and the Gaussian kernel parameter is 110. In addition, based on the information of the international spring and summer women's fashion color definitions released by the International Color Council from 2017 to 2022, the study collects and organizes nearly 300 colors and uses the SWK-FCM clustering algorithm to find ten clustering centers. After data processing, the classification results and the contents of the classification intervals are shown in Table I.

TABLE I. TEN CLUSTERING CENTERS AND HUE INTERVALS OF HUE VALUES IN PANTONE COLOR SPACE

Hue classification	Name	Cluster center	Chromatic value range
1	Green	2	[1, 4].
2	Greenish yellow	6	[5, 8].
3	Yellow	10	[9, 11].
4	Yellow-red	13	[12, 15].
5	Red	17	[16, 21]
6	Red-purple	26	[22, 30]
7	Purple	39	[31, 40].
8	Purplish blue	44	[41, 48]
9	Blue	53	[49, 57].
10	Turquoise	61	[58, 64].

In Table I, a total of ten cluster centers were obtained in this study. The corresponding colors are mainly green, yellow, red, purple, blue, and five intersecting colors. On this basis, the classification quality assessment criteria are selected as classification coefficient (E) and classification entropy (F). The experimental clothing images are selected as 5% Gaussian noise and 10% pretzel noise. The comparison results of the performance parameters of the three algorithms classification criteria simulation are shown in Fig. 5.

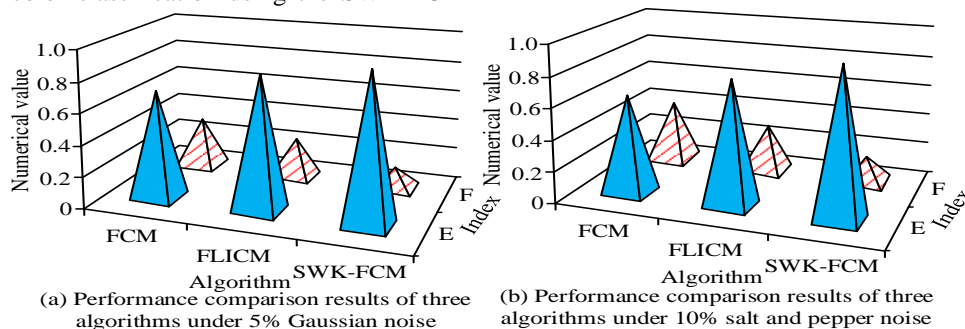


Fig. 5. Comparison results of performance parameters of clothing color hue tested by three algorithm classification standards.

Fig. 5(a) shows the comparison results of the algorithms under 5% Gaussian noise. Fig. 5(b) shows the comparison results of the algorithms under 10% pretzel noise. From Fig. 5(a), the classification coefficient of SWK-FCM clustering

algorithm is 0.9553, which is higher than 0.7063 of FCM and 0.8598 of FLICM. Meanwhile, the classification entropy of SWK-FCM clustering algorithm is 0.1380, which is lower than 0.3324 of FCM and 0.2623 of FLICM. At the same time,



the classification entropy of SWK-FCM clustering algorithm is 0.1837, which is lower than that of FCM (0.3598) and FLICM (0.3157). This indicates that the SWK-FCM clustering algorithm has better classification quality and more robustness. To further verify the result, the study applies it to

the actual apparel color classification, in which three colors, green, yellow-red and purple, are randomly selected under ten clustering centers. The three colors are used as the clothing classification images for the experiment. The comparison results are shown in Fig. 6.

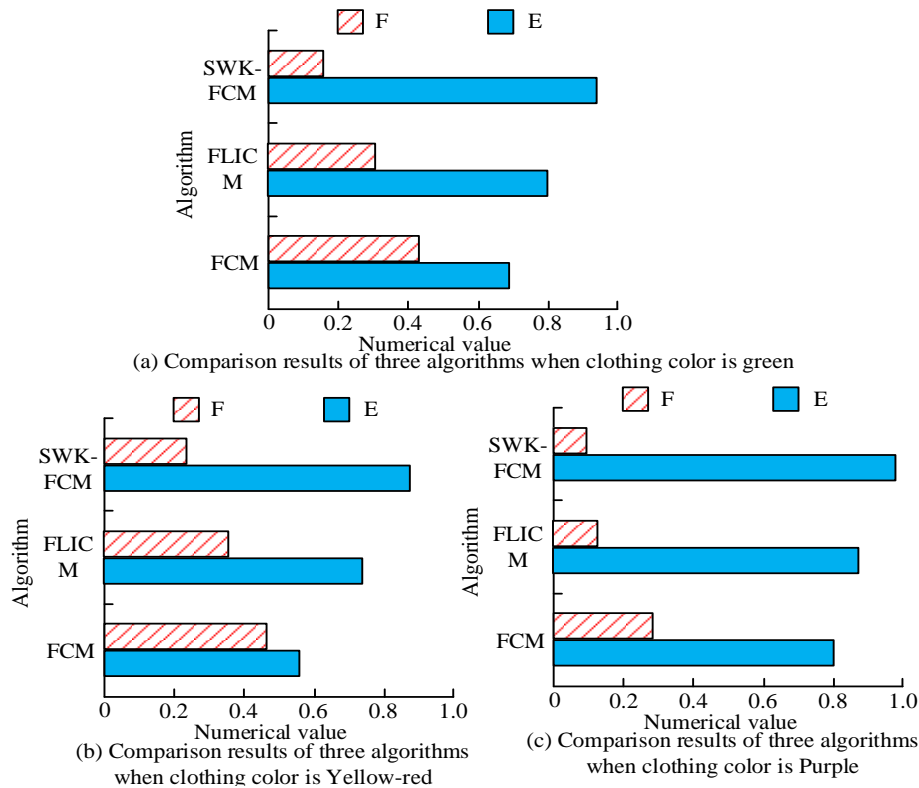


Fig. 6. Performance comparison results of three algorithms for clothing classification images.

Fig. 6(a) shows the comparison results of the three algorithms when the garment hue is green. Fig. 6(b) shows the comparison results of the three algorithms when the garment hue is yellow-red. Fig. 6(c) shows the comparison results of the three algorithms when the garment hue is purple. From Fig. 6(a), the classification coefficient of SWK-FCM clustering algorithm is 0.9388, which is much higher than that of the comparison algorithm. The classification entropy is 0.1540, which is significantly lower than that of the comparison algorithm. From Fig. 6(b), the classification coefficient of SWK-FCM clustering algorithm is 0.8967, which is higher than the comparison algorithm. The classification entropy is 0.2637, which is lower than the comparison algorithm. In Fig. 6(c), the classification coefficient of SWK-FCM clustering algorithm is 0.9778, which is also higher than that of the comparison algorithm. The classification entropy is 0.0946, which is lower than that of the comparison algorithm. Therefore, the SWK-FCM clustering algorithm is less susceptible to interference from noise and has higher classification accuracy in the actual clothing color classification. Therefore, the two prediction models constructed on the basis of SWK-FCM are studied to possess better performance in theory. To verify the performance of the GM (1, 1) prediction model and the BP neural network prediction model, the study analyzes the prediction results of both of them separately. In the

experiments, the study compares the absolute error, relative error, and Mean Square Error (MSE), which are denoted by e, d, and m, respectively. The results are shown in Fig. 7.

In Fig. 7, G (2022) denotes the original value in 2022. P (2022) denotes the predicted value in 2022. e denotes the absolute error; d denotes the relative error. m denotes the mean square error. Comprehensive Fig. 7 shows that the MSE values of the GM (1, 1) model color prediction using the final case information are very low, at 0.000004~0.001296. Its MSE can be as small as  $10^{-6}$  magnitudes, and the overall MSE is 0.00028, which is close to the magnitude of  $10^{-4}$ . This indicates that the GM (1, 1) model has higher prediction ability and higher accuracy rate. Meanwhile, based on the comparison between the prediction and the real curve in 2022, the average relative error of the popular colors in 2022 is 14.51%, which is very close to the actual value, except for the relatively large relative error of individual colors. This indicates that the GM (1,1) model using the final case information has a stronger trend inference ability. In addition, using the constructed BP neural network prediction model to determine case information, a time series is created for four consecutive years of raw color data from 2019 to 2022. It serves as the input feature vectors a, B, C, and D., The predicted results obtained through continuous training are shown in Fig. 8.

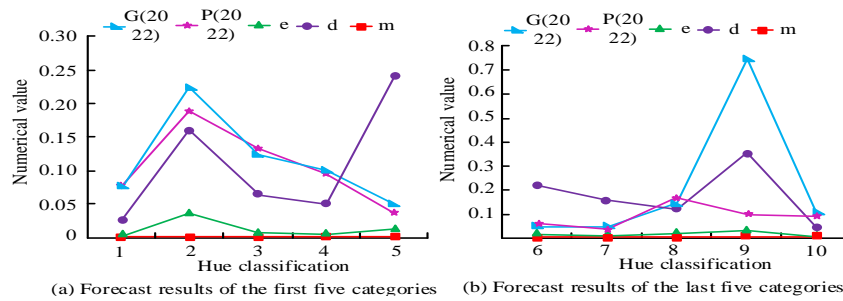


Fig. 7. Prediction results of SWK-FCM and GM (1,1) based on final information.

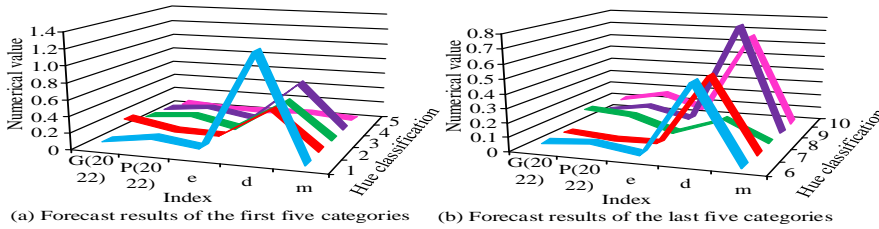


Fig. 8. BP neural network prediction results based on final information.

Comprehensive Fig. 8 shows that the MSE value of the BP neural network prediction model with 4-7-1 network structure for the popular color hues is between 0.000529 and 0.011025. While the MSE of the four colors, red, yellow-green, green, green-blue, and purple is of the order of  $10^{-4}$ , and the overall MSE is 0.0036, which is about the order of magnitude of  $10^{-2}$ . In summary, the prediction model using SEW-FCM and BP neural network is not as accurate as the GM (1,1) prediction model. But it can improve the prediction accuracy in comparison with other models. On this basis, to more comprehensively analyze the high accuracy of the prediction model of clothing fashion color built on the basis of SWK-FCM clustering algorithm, the study compares the two models built using the final case information with the prediction models built on the basis of traditional classification methods. The four models are represented by GM (1,1) (F), BP(F), GM and BP. Among them, the BP neural network uses relu as the excitation function, with a maximum number of iterations of 1000, 4 input feature vectors, 4 input layer nodes, 1 hidden node, and 1 output layer node. The minimum error of Xunyu is less than  $10^{-4}$ . The results are shown in Fig. 9.

The MSE value of the GM (1, 1) prediction model using the final case information is 0.00028, while the MSE value of the BP neural network prediction model using the final case information is 0.0036. Both of them are much lower than the 0.0071 of the GM model. The GM (1, 1) prediction model is

ideal, indicating that it can better predict the trend of popular colors.

Finally, to further validate the performance of the proposed SWK-FCM algorithm, a clustering algorithm using graph theory (a), a fuzzy kohlen clustering network algorithm (b), and a fuzzy covariance learning vector quantization algorithm (c) are introduced and compared with FCN (d) and SWK-FCM (e). Among them, the comparison indicators include accuracy, accuracy, recall, F1 value, and area under the curve, represented by 1-5. The results are shown in Table II.

From Table II, the algorithm proposed in the study has values of 98.95%, 98.65%, 99.01%, 99.63%, and 98.73% on indicators 1-5, respectively, which are higher than the comparative algorithms. This indicates the effectiveness and high performance of the algorithm in actual classification.

TABLE II. COMPARISON RESULTS OF DIFFERENT ALGORITHMS AND INDICATORS

-	A	B	C	D	E
1	92.41%	90.91%	95.45%	90.91%	98.95%
2	88.45%	92.34%	96.45%	89.65%	98.65%
3	90.21%	95.45%	97.11%	90.01%	99.01%
4	89.00%	93.16%	98.21%	89.51%	99.63%
5	82.06%	91.51%	94.63%	79.41%	98.73%

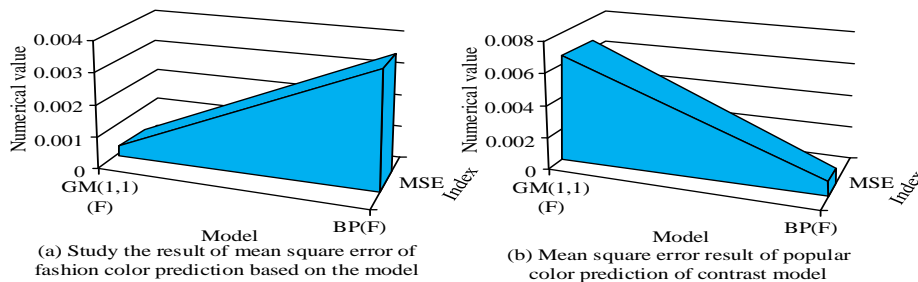


Fig. 9. Comparison results of popular color hue prediction based on final information with other models.

## V. CONCLUSION

To achieve the prediction of the actual cutting-edge fashion information of current fashion colors, the study obtained the SWK-FCM clustering algorithm by improving the traditional FCM clustering algorithm. Meanwhile, the prediction model of GM (1, 1) and BP neural network was constructed on the basis of this clustering algorithm combined with the final case information. The performance of SWK-FCM clustering algorithm and the effect of the two prediction models were analyzed by experiments. The experimental results showed that the classification coefficients of SWK-FCM clustering algorithm under different conditions were 0.9553 and 0.9258, and the classification entropies were 0.1380 and 0.1837, respectively. The same results were presented in the actual clothing color classification. In the GM (1, 1) prediction model experiments, the MSE of the GM (1, 1) model color prediction using definite case information was in the range of 0.000004 to 0.001296, and its MSE could be as small as  $10^{-6}$  magnitudes. In the experiments of the BP neural network prediction model, the MSE values ranged from 0.000529 to 0.011025 for the prediction of popular color shades, and the overall MSE was 0.0036, which reaches the magnitude of  $10^{-2}$ . Among the four models compared, the mean square error value of the GM (1, 1) prediction model using the final case information was 0.00028, which is lower than the other compared models.

Overall, the SWK-FCM clustering algorithm has better classification quality, is less susceptible to noise interference in actual color classification, and can automatically classify color phases. In theory, it can also be applied to other color spaces. Among the prediction models constructed on its basis, GM (1,1) has a better prediction effect, and both can provide corresponding guidance and assistance to stakeholders in the clothing ecosystem. However, the amount of popular color finalization information selected for the study is too small, which increases the difficulty of the study. Therefore, it is necessary to increase the amount of data in the future. At the same time, the study only analyzed the hue among the three attributes of color. In the future, a comprehensive prediction of popular colors should be made based on the saturation and purity of colors.

## REFERENCES

- [1] Kodžoman D, Hladnik A, Pavko Čuden A, Čok, V. Exploring color attractiveness and its relevance to fashion. *Color Research & Application*, 2022, 47(1): 182-193.
- [2] Papadopoulos S I, Koutlis C, Papadopoulos S, Kompatsiaris I. Multimodal Quasi-AutoRegression: Forecasting the visual popularity of new fashion products. *International Journal of Multimedia Information Retrieval*, 2022, 11(4): 717-729.
- [3] Zhu D, Lai X, Rau P L P. Recognition and analysis of kawaii style for fashion clothing through deep learning. *Human-Intelligent Systems Integration*, 2022, 4(1-2): 11-22.
- [4] Zhu H, Lv J, Hu Y, Liu C, Guo H. Application of K-means algorithm in Yi clothing color//International Conference on Internet of Things and Machine Learning (IoTML 2021). *SPIE*, 2022, 12174: 236-240.
- [5] Garcia C C. Fashion forecasting: an overview from material culture to industry. *Journal of Fashion Marketing and Management: An International Journal*, 2022, 26(3): 436-451.
- [6] Khaydarova L, Sarvinoz N. Translation in Fashion and the art of dressing. *INTERNATIONAL JOURNAL OF SOCIAL SCIENCE & INTERDISCIPLINARY RESEARCH ISSN: 2277-3630 Impact factor: 7.429*, 2022, 11(02): 64-66.
- [7] Cheng Y, Zhou F, Zhao Y. Measurement and prediction of the international reputation of Chinese women's apparel and accessories. *Textile Research Journal*, 2023, 93(1-2): 194-205.
- [8] Zhou Z, Deng W, Wang Y, Zhu Z. Classification of clothing images based on a parallel convolutional neural network and random vector functional link optimized by the grasshopper optimization algorithm. *Textile Research Journal*, 2022, 92(9-10): 1415-1428.
- [9] Mushtaque S, Siddiqui A A, Wasim M. An ontology based approach to search woman clothing from Pakistan's top clothing brands. *KIET Journal of Computing and Information Sciences*, 2022, 5(1): 37-47.
- [10] Wu D, Li Z, Zhou J, Gan J, Gao W, Li H. Clothing attribute recognition via a holistic relation network. *International Journal of Intelligent Systems*, 2022, 37(9): 6201-6220.
- [11] Dai Y, Chen Y, Gu W, Tan Y, Liu X. Color identification method for fashion runway images: An experimental study. *Color Research & Application*, 2022, 47(5): 1163-1176.
- [12] Han A, Kim J, Ahn J. Color Trend Analysis using Machine Learning with Fashion Collection Images. *Clothing and Textiles Research Journal*, 2022, 40(4): 308-324.
- [13] Mau T N, Inoguchi Y, Huynh V N. A novel cluster prediction approach based on locality-sensitive hashing for fuzzy clustering of categorical data. *IEEE Access*, 2022, 10: 34196-34206.
- [14] Mersch B, Chen X, Vizzo I, Nunes L, Behley J, Stachniss C. Receding moving object segmentation in 3d lidar data using sparse 4d convolutions. *IEEE Robotics and Automation Letters*, 2022, 7(3): 7503-7510.
- [15] Shen Y, Sun Y, Li X, Eberhard A, Ernst A. Enhancing column generation by a machine-learning-based pricing heuristic for graph coloring//Proceedings of the AAAI Conference on Artificial Intelligence. 2022, 36(9): 9926-9934.
- [16] Buchel S, Hebinck A, Lavanga M, Loorbach D. Disrupting the status quo: a sustainability transitions analysis of the fashion system. *Sustainability: Science, Practice and Policy*, 2022, 18(1): 231-246.
- [17] Wang W, Liu Y, Song F, Wang Y. Color trend prediction method based on genetic algorithm and extreme learning machine. *Color Research & Application*, 2022, 47(4): 942-952.
- [18] Jiang R, Wang Y, Li J. Prediction of clothing comfort sensation with different activities based on fuzzy comprehensive evaluation of variable weight. *Textile Research Journal*, 2022, 92(21-22): 3956-3972.
- [19] Majhi R, Sugasi R P. A machine-learning approach for classifying Indian internet shoppers. *Applied Marketing Analytics*, 2022, 7(3): 288-298.
- [20] Tian X, Wu W, Ma X, Zhang P. A new information priority accumulated grey model with hyperbolic sinusoidal term and its application. *International Journal of Grey Systems*, 2021, 1(2): 5-19.
- [21] Yousuf M U, Al - Bahadly I, Avci E. Wind speed prediction for small sample dataset using hybrid first - order accumulated generating operation - based double exponential smoothing model. *Energy Science & Engineering*, 2022, 10(3): 726-739.
- [22] Zhang J, Qin Y, Zhang X, Che G, Sun X, Duo H. Application of non-equidistant GM (1, 1) model based on the fractional-order accumulation in building settlement monitoring. *Journal of Intelligent & Fuzzy Systems*, 2022, 42(3): 1559-1573.
- [23] Liu K, Wu H, Zhu C, Wang J, Zeng X, Tao X, Bruniaux P. An evaluation of garment fit to improve customer body fit of fashion design clothing. *The International Journal of Advanced Manufacturing Technology*, 2022, 120(3-4): 2685-2699.
- [24] Xue J, Nie F, Wang R, Li X. Iteratively Reweighted Algorithm for Fuzzy  $\$ c \$$ -Means. *IEEE Transactions on Fuzzy Systems*, 2022, 30(10): 4310-4321.

# Next-Generation Intrusion Detection and Prevention System Performance in Distributed Big Data Network Security Architectures

Michael Hart<sup>1</sup>, Rushit Dave<sup>2</sup>, Eric Richardson<sup>3</sup>

College of Science, Engineering, & Technology, Minnesota State University, Mankato, United States<sup>1,2</sup>  
College of Health and Human Services, University of North Carolina Wilmington, United States<sup>3</sup>

**Abstract**—Big data systems are expanding to support the rapidly growing needs of massive scale data analytics. To safeguard user data, the design and placement of cybersecurity systems is also evolving as organizations to increase their big data portfolios. One of several challenges presented by these changes is benchmarking real-time big data systems that use different network security architectures. This work introduces an eight-step benchmark process to evaluate big data systems in varying architectural environments. The benchmark is tested on real-time big data systems running in perimeter-based and perimeter-less network environments. Findings show that marginal I/O differences exist on distributed file systems between network architectures. However, during various types of cyber incidents such as distributed denial of service (DDoS) attacks, certain security architectures like zero trust require more system resources than perimeter-based architectures. Results illustrate the need to broaden research on optimal benchmarking and security approaches for massive scale distributed computing systems.

**Keywords**—Big data systems; zero trust architecture; benchmarking; distributed denial of service attacks

## I. INTRODUCTION

Big data systems are unified environments designed for massive-scale data analytics. Systems capable of handling large amounts of data are becoming more important as the volume of data created and communicated over the Internet increases [1]. Cybersecurity systems play an important role in ensuring the large quantities of data on the Internet remains safe. One dimension of several necessary to accomplish the latter are next-generation security devices. Intrusion detection and prevention systems (IDPSs) properly manage data accessibility, privacy, and safety. IDPS algorithms are able to identify cyber threats using several mechanisms. This includes using prior information from previous attacks, anomalies in network packets [1], and machine learning [2].

As big data systems become more common, their roles will continue to expand. This includes the capability to analyze and detect information security vulnerabilities at scale. For example, several big data frameworks exist that discover distributed denial of service (DDoS) attacks [3]. This expansion of roles offers many exciting opportunities for organizations. However, as the use of big data systems grows, the capability of attackers to leverage associated parallel computing power for nefarious reasons also increases [3]. A

systematic review of 32 papers pertaining to securing big data found that a critical need in future research is building more secure big data infrastructure [4]. Contributing to the latter objective, the researchers demonstrate how varying network architectures impact the security and performance of big data systems.

Organization of the paper is as follows. Section II reviews literature on intrusion detection and prevention methods for big data systems. Section III outlines the research design and methodologies used to test perimeter-based security and perimeter-less security applied to a big data system environment. Section IV describes the research results. Section V concludes the study by discussing the limitations and future outlook.

## II. LITERATURE REVIEW

Work is necessary to optimize both the information security and performance of distributed systems. Today, several open-source big data frameworks provide remarkable potential for solving challenging data science and related problems by leveraging powerful parallel and distributed data processing. However, securing these systems often carries performance penalties. The review of literature that follows explores research on the impact of various IT infrastructure security strategies and their influence on big data environments. It begins by reviewing comprehensive surveys most closely related to information security and big data systems.

### A. Surveys of Big Data and Intrusion Detection

Previous systematic reviews of literature focused on information security and big data provide a vast array of objectives. A prominent theme is using deep learning [1] and machine learning [2] to assist in detecting or preventing cybersecurity attacks. This line of research often utilizes deep learning or machine learning algorithms for near real-time data protection.

A recent and well cited comprehensive survey in [1] evaluates how deep learning is used for intrusion detection systems in the cybersecurity domain. It found notable contrasts between machine learning approaches in cybersecurity and deep learning. Conventional machine learning approaches utilized in cybersecurity were classified by approaches such as artificial neural networks (ANNs), Bayesian networks, decision trees, fuzzy logic, k-means clustering, k-nearest neighbor (kNN) algorithm, and support vector machines (SVMs). The

survey centered on deep learning focal intrusion detection methods that included autoencoders (AEs), convolutional neural networks (CNNs), deep belief networks (DBNs), generative adversarial networks (GANs), and long short-term memory (LSTM) recurrent neural networks [1].

AEs, DBNs, and GANs were highlighted in [1] for their unsupervised learning strengths. In the absence of gradient estimation, AEs can use gradient descent to train data. A strength of LSTM is its capabilities in analyzing time-series data. CNNs do not need as much data processing prior to evaluation as certain algorithms and is able to classify cyber-attacks using multiple characteristics well. Combined, the survey of literature finds that AEs, CNNs, DBNs, GANs, and LSTM networks each have potential to improve intrusion detection methods. Furthermore, the survey [1] outlined the importance of dataset reliability when evaluating deep learning intrusion detection effectiveness. Variance in cybersecurity attack datasets can introduce model bias when comparing multiple deep learning methods. Thus, any biases in attack datasets or data from live systems could increase spurious results [1].

A subsequent theme in the literature concentrates on cybersecurity and privacy prevention in big data applications. While this research again employs various data science methods to detect or prevent data breaches, it also illustrates how big data techniques can prevent information privacy issues. Research in [4] led to a proposed model for enhancing information privacy. The model highlights people, organizations, society, and government roles. It leverages IDS, IPS, and encryption as its primary techniques to prevent data breaches [4].

### B. Big Data Architectures and Information Security

As big data evolves, the supporting infrastructures will require proper encryption, intrusion detection, and intrusion prevention. Changing architectures within computer networks, messaging techniques, and undefined communication methods introduce numerous challenges. In a 2014 study Mitchel and Chen [5] recognized this paradigm. Their emphasis on cyber-physical systems (CPS) ranging from smart grids to unmanned aircraft systems led to the classification of four primary intrusion detection categories. These include legacy technologies, attack sophistication, closed control loops, and physical process monitoring. Each of the latter is narrow concepts as they relate to the broader field of intrusion detection, underlying the unique customization of IDSs for cyber-physical systems [5].

Three years later Zarpelo et al. [6] outlined a similar but distinct paradigm; intrusion detection focal to the Internet of things (IoT). The researchers stated that IoT has similar information security matters as the Internet, cloud services, and wireless sensor networks (WSNs). Despite similarities, IoT information security approaches are distinct, according to the authors due to concepts such as data sharing between users, the volume of interconnected objects, and the amount of computational power of the associated devices. Like cyber-physical systems, IoT presents diverse challenges to the design of intrusion detection systems [6].

Designing secure cloud computing environments poses several novel problems at multiple infrastructure layers. As an example, cloud resources can be leased by numerous vendors focused on varying as-a-service models such as infrastructure as a service (IaaS), platform as a service (PaaS), and/or software as a service (SaaS). Multi-cloud applications rely upon the seamless integration of cloud resources from providers focused on one or many as-a-service types, which continue to expand. In Casola et al. [7] a model is outlined for designing, creating, and implementing multi-cloud applications. The flexible approach accounts for varying as-a-service components. Security-by-design is a primary objective of the process lifecycle between the functional design of multi-cloud applications and the security design. The functional design phase defines the application logic, interconnections of services, and resource requirements. In the security design phase, each cloud element is assessed in terms of security risks and security needs. Security policies and controls are designed based on the latter requirements. Similar to CPS [5] and IoT [6], the multi-cloud application model is a subsequent example of how information security solutions play a prominent role due to the systems' distinct architectural and infrastructure layers.

Securing big data environments or leveraging associated techniques like machine learning to enhance information security intertwines numerous fields include but not limited to CPS, IoT, and cloud computing. Like big data systems, CPS requires cybersecurity protection [8] of private data [9]. Big data, IoT, and CPS often overlap through the ad hoc interfaces of systems such as smart vehicles, buildings, factories, transportation systems, and grids [10]. As a vulnerable attack surface, IoT advances the need for intelligent information security.

Machine learning [11], including ensemble intrusion detection [12], and IDS design [13] are proposed techniques to mitigate malicious cybersecurity attacks. Due in part to porous attack surfaces in cloud centric big data, IDSs may require collaborative frameworks [14]. In [15], fuzzy c means cluster (FCM) and support vector machine (SVM) were proposed as a collaborative technique for IDS detection rates. Compared to other mechanisms, the proposed hybrid FCM-SVM showed lower false alarm ratios and higher detection accuracy [15]. Furthermore, [16] illuminates the need for scaling IDS detection algorithms using the resources of parallel computing in the cloud.

In [17] the researchers propose the BigCloud security-by-design framework. The framework draws from the need to integrate big data security into the system development lifecycle. Its primary cloud application domain is focal to infrastructure as a service. It notes IaaS as one of the faster growing as-a-service options for big data. The model helps design and enforce secure authentication, authorization, data auditability, availability, confidentiality, integrity, and privacy. However, its IaaS concentration could provide greater benefits to as-a-service components specific to host operating systems, hypervisors, networking, and hardware [17]. Similar to IaaS, the evolution of serverless platforms and Function-as-a-service (FaaS) applications requires careful security design to overcome security threats that new services often suffer [18].

While distinct, CPS, IoT, cloud computing, and big data are merely a few examples of why designing intrusion detection and prevention systems remains highly elastic in modern computational architectures. As the information technology landscape changes, information security bends to meet the evolving needs of the complete environment. To conclude the literature review, the authors will outline several relevant studies introducing potential solutions to design stronger information security controls for big data systems.

### C. Encryption

An ongoing challenge in distributed big data systems is securing communication between multiple systems operating across various computer networks. Apache Hadoop and Apache Spark are examples of big data frameworks that present several opportunities for attackers to access the data they facilitate. Central to big data frameworks is the ability to use parallel processing to analyze massive amounts of data. MapReduce is one of many programming paradigms that leverages Hadoop to extract valuable knowledge from large volumes of data. However, like most application or service modules within big data frameworks, MapReduce highlights the vast attack vectors that exist in distributed big data systems. MapReduce examples in literature include side channel attacks [19], job composition attacks [20], and malicious worker compromises in the form of distributed denial-of-service (DDoS) or replay attacks [21], Eaves dropping and data tampering [22]. Encryption is a primary countermeasure to secure transmissions and prevent data leaks between big data servers [19].

A primary objective in addressing cybersecurity attacks on parallel processing services is identifying and preventing leaks that often occur during data transmission between distributed worker nodes, also referred to as DataNodes in Apache Hadoop. These unique yet integrated servers work in parallel to complete MapReduce jobs. Often in Hadoop, data is stored and retrieved from the Hadoop Distributed File System (HDFS). In [19] side-channel attacks are addressed that can occur between MapReduce workers that utilize HDFS for data storage. These types of cybersecurity attacks can target worker nodes to extract valuable information pertaining to MapReduce jobs such as the amount of packet bandwidth. This further contributes to successful pattern attacks. The authors proposed a solution to this vulnerability labeled Strong Shuffle that enforces strong data hiding between workers [19]. In contrast to alternative countermeasures such as correlation hiding in [20], Strong Shuffle avoids leaking the number of records accepted by each reducer during MapReduce runtime. Secure plaintext communications is a function of semantically secure encryption in the Strong Shuffle solution [19].

In [19] data communicated between Hadoop DataNodes and stored in HDFS is encrypted with semantically secure AES-128-GCM encryption. Although the latter helps prevent clear text leakage between MapReduce jobs in Hadoop, encryption in big data environments has limitations. For example, encrypted databases can still reveal certain information during operations that include table queries. Deterministic encryption and order-preserving encryption can leak the equality relationship and the order between records. One proposed solution is semantically secure encryption. In

[23] the authors propose a semantically secure database system named Arx. Alternative to order-preserving encryption, semantic security within Arx only allows an attacker to extract order relationships and frequency of the direct database query in use in contrast to the entire database. The authors note that worst-case attackers would gain as much information from a data leak as deterministic or order-preserving encryption over time [23]. While methods such as encryption and authentication help with cross-node data leaks, they do not prevent other attacks, such as DDoS and passive network eavesdropping [21]. A subsequent countermeasure is the effective design and implementation of intrusion detection and prevention systems [14].

### D. Next-Generation Security and Big Data Systems

Next-generation security at a high level can detect and prevent malicious cybersecurity attacks. Much of the literature focuses on identifying malicious network packets in real-time. The comprehensive survey in [24] reviews how modern data mining techniques are evolving to meet real-time detection needs. The review classifies intrusion detection systems by architecture, implementation, and detection methods. Detection methods are categorized as anomaly-based, signature based, and hybrids. Signature based methods or misuse often rely upon a database that defines patterns or existing malicious attack signatures. Anomaly detection can detect non-normal network traffic behavior that has yet to be defined in a signature database. Data mining methods including supervised, unsupervised, and hybrid learning are being used to improve anomaly-based intrusion detection systems [24].

While supervised, unsupervised, and hybrid learning IDS research continues to progress [24], the ongoing need to improve existing big data implementations remains. In several systematic literature reviews [1, 2, 3, 24], IDSs are known to have limitations that contradict the performance benefits of parallel processing and distributed computing. For example, large signature based systems drain CPU and memory resources [24]. While researchers continue to advance areas of intrusion detection such as packet anomalies and encryption, only a few studies are advancing security by design and its effects on varying big data architectures [1]. To address this need, the authors of this study designed a distributed big data system over a wide area network to explore the performance of distributed nodes under different network traffic loads.

## III. METHODS

This research methodology follows the design science approach in [25] and [26]. Design science is based on a scientific framework for IT research. As March and Smith [25] outline, IT research should consider natural and design science as a method to build and evaluate tangible objects. Within this philosophy, objects often have outputs in the form of models or instantiations. Instantiations associate with new artifacts in the design science methodology and the understanding of the artifact in its environment [25]. IT artifacts can be realized in many forms such as through the design of an object that helps solve business problems [26].

### A. Organizational Problem

Central to the organizational problem in this study is the need to architect a real-world or simulated big data environment that generates important inputs and outputs. In the case of this study, several architectural layers require design, configuration, benchmarking, and evaluation that accurately represent industry big data system implementations. These research activities could establish a more mature model for IDPS placement in evolving network architectures. Design science methods guide the latter activities [26].

Big data clusters can have thousands of nodes. Attempting to secure individual servers poses several issues ranging from significant costs to lost computational resources. Important to the artifact design process is the creation of an IDS and IPS testing environment that results in minimal disruption to existing big data infrastructures. Additionally, the authors constructed an experimental setup similar to several local small business environments that are readily available, relatively inexpensive, and relevant to a broad audience. Therefore, the testing environment is limited to several small commodity virtual machines (VMs) operating in physically distanced data centers. The authors will briefly outline the network architecture, hardware, software used in the experimental environment.

### B. Network Architecture

Fig. 1 depicts the baseline network architecture used in this study. The experimental network emulates a small to medium-sized business with a 200 Mbps dedicated lease line between four distinct physical locations. Connections are 1 Gbps copper from the demarcation point to the LAN nodes. Each server is connected to layer 2 switches followed by a layer 3 Cisco Systems enterprise class router.

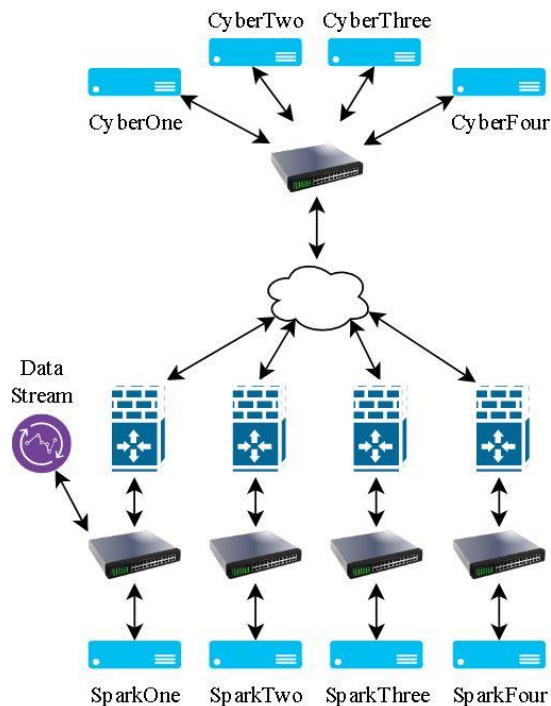


Fig. 1. Perimeter-based security network architecture.

The cybersecurity servers labeled “CyberOne” to “CyberFour” illustrate the systems used to attack the big data cluster. The big data cluster includes four servers labeled “SparkOne” to “SparkFour.” One streaming server is depicted as the data stream located in the same local area network (LAN) as SparkOne. Four intrusion detection and prevention systems are situated between each big data server and its extrinsic networks.

### C. Hardware

The big data servers run on parallel Dell hardware [27]. The hardware is manufactured on the same date and shipped in the same container. The testing server used the same single Intel CPU with 16 logical cores and 32 GBs of physical random-access memory. The baseline Intel CPU benchmark average results from the PassMark version 10 performance test [29] are 2,799 MOps per second for a single thread and 5,443 megabytes per second for data encryption.

Cisco RV series routers with integrated firewalls exist between each Apache Spark node and the external network. Cisco Firmware 1.0.3.55 is in use with the default firewall ruleset. The authors added customized rules that allow the internal LAN IP addresses to communicate on the necessary Apache HDFS and Spark ports. Subsequent ports are blocked [28].

### D. Big Data Systems

Each big data server and streaming server used equivalent software and versions. Systems ran on the Ubuntu server 20.04.3 LTS operating system. Installed software included Java 11, Python 3.8, Apache Hadoop 3.2, and Apache Spark 3.2. The big data environment is comprised of five servers. This includes one primary cluster manager labeled *SparkOne* and three secondary work nodes labeled *SparkTwo*, *SparkThree*, and *SparkFour*. Apache Spark is tuned using optimal parameters such as those specified in [30] and [31]. HDFS disks are balanced between nodes with DFS replicating three blocks. The data stream denotes the independent Spark streaming instance.

SparkOne is the primary node in the testing environment used in this study. It is comprised of the driver program. The driver program executes the big data application’s main() class and generates the SparkContext [32]. SparkContext is capable of using various big data resource managers. Tests in this study use Yet Another Resource Negotiator (YARN) as the distributed cluster manager [33].

SparkContext helps communicate application jobs containing code in various forms such as Python and JAR files to the executors on the worker or secondary nodes in the cluster. YARN has two primary high-level components labeled the NodeManager and ResourceManager. Secondary nodes in a big data cluster managed by YARN each have a NodeManager. Its function is to manage containers on each server. Containers encompass resources such as network, disk, CPU, and memory. These are allocated properly to facilitate task execution. The YARN ResourceManager consists of the ApplicationsManager and the Scheduler. While the Scheduler determines the necessary resources for each application the

ApplicationsManager identifies which container the application will use and subsequently monitors their task execution [33].

Apache Spark and HDFS replicate between three secondary big data servers. The secondary or worker nodes labeled SparkTwo, SparkThree, and SparkFour contain executor processes. An executor process remains throughout the runtime of tasks that each worker is allocated by the cluster manager. Every application receives its own executor process and/or processes as necessary. The driver program on SparkOne is configured to listen for executor process communications from the secondary nodes until the job is completed. Per Apache Spark documentation in [32], when possible, the driver program should be on the same local area network as the worker nodes due to the latter communication. In the experimental network design, the worker nodes are physically distanced. Therefore, Spark is optimized to open local remote procedure calls on the worker LANs [32].

#### E. Attack Systems

Although the cybersecurity servers ran on the same hardware as the big data servers, they used different software. CyberOne, CyberTwo, CyberThree, and CyberFour each delineate a server used to carry out cyber-attacks on the big data cluster. The software includes the Kali Linux operating system running the 5.14 kernel. Kali Linux is an open-source operating system based on Debian Linux. It is designed for numerous information security objectives such as reverse engineering, forensics, pen testing, and research [34].

#### F. Intrusion Detection and Prevention Systems

Consistent with Fig. 1, the baseline IDS and IPS systems are located between the cyber-attack and big data systems. Regardless, the authors manipulate the placement of these systems throughout each experimentation. As a simulated construct in the research methodology, the authors propose that IDS and IPS architecture placement predicts data streaming performance between worker nodes. Performance evaluation of this potential construct is an important step toward advancing a future IDPS placement framework for physically distanced big data systems.

The authors implemented Snort and Suricata, two popular open-source IDS and IPS systems. Snort is developed by Cisco Systems. It serves as a leading intrusion detection engine and rule set for Cisco next-generation firewalls and IPSs. Its mechanisms for detecting and preventing security threats continue to evolve. However, a fundamental capability during this writing is the formation of rules. In contrast to traditional methods such as signature-based detection, rules focus on vulnerability detection [35]. Suricata is developed by the Open Information Security Foundation (OISF). Similar to Snort, Suricata can use rules to detect and block cyber-attacks [36].

Version 2.9.7 of Snort ran with libpcap version 1.9.1 and version 8.39 of the payload detection rules. Suricata testing uses version 6.0.6 with the emerging threats open ruleset. The authors customized the latter default Snort and Suricata rulesets to secure the distributed nodes. The rulesets are parallel in count and type (e.g. alert, drop) to control significant variations in resource contention. Suricata and Snort use the same rules in the tests, except for minor incompatibilities. Where

incompatible, the rules are adjusted to perform the same action in both IDSs at parallel throughput rates.

Snort and Suricata run on the same server hardware and operating systems as the big data servers. A second NIC allows the servers to act as gateways between trusted and untrusted networks. The servers communicate between the local area networks using Transport Layer Security (TLS) and Secure Shell (SSH) Protocols. Ubuntu server 20.04.3 LTS is configured using OpenSSH version 8.2 and OpenSSL version 1.1.1.

#### G. Benchmarks

The authors developed custom benchmarks to identify how big data clusters perform under various IDS physically distanced network architectures. The benchmarks perform two significant network load functions, 1) streaming unstructured data to the Spark big data cluster and 2) flooding the Spark nodes via DDoS attacks. Network and system benchmarking uses version 16m of the nmon source code to measure network performance. Originally developed by IBM, nmon is an open-source Linux project that monitors system resource utilization. Performance metrics include CPU, disk, memory, and networking [37].

The authors follow the design science methodology [25] to design and implement an IDS placement experiment for physically distanced big data systems. Next, the authors construct a series of tests to determine how IDS locations influence real-world distributed worker nodes.

## IV. RESULTS

Each of the tests followed an eight-step process, 1) network architecture is determined and implemented, 2) IDPS locations are identified and configured, 3) IDPS customized rulesets are implemented, 4) the big data system cluster is started and tested as operational, 5) data streams to the cluster are invoked, 6) DDoS attacks are executed, 7) the benchmarks are run, and 8) the researchers maintain and monitor the testing environment for anomalies. Each of the tests was repeated three times to ensure saturation existed in the results.

#### A. Test 1 Perimeter-Based Security Results

Fig. 1 illustrates the IDPS placement location for the first test. The cloud represents the leased line between the geographical sites. Below the cloud icon is the selected IDPS solution followed by the Apache Spark cluster. Network architecture in the first test follows Cisco Systems' best practices for a collapsed data center and LAN core [38]. Within this design, a hardware-based IDPS is situated between the public untrusted and private trusted networks. Test one includes a traditional perimeter Cisco Systems IDPS. Individual Spark nodes are networked in a single VLAN connected through the collapsed core.

In contrast to the network architecture in Fig. 1, CyberOne through CyberFour servers are not deployed for tests 1-3. In each of these tests, typical network traffic is present void of any DDoS attacks.

Benchmark metrics are specific to the big data systems unless otherwise specified. During the data stream, HDFS is



writing 128 MB blocks to disk on all three Spark worker nodes at a constant rate. Inconsequential wait time exists on disk reads and writes. Average CPU utilization per thread or “CPU%” on the big data worker nodes is 4.3% during the first test. The average time a process waits for an input-output (I/O) to complete or “wait%” is 0.3. The average number of processor context switches per second is 1,728, identified as “PWps” hereafter.

The authors measured network performance between each of the Spark nodes using four metrics. Metrics are captured on the worker node network interface cards. The first performance variable measures the average number of all network packet reads per second (APRps). The second variable captures the average number of all network packet writes per second (APWps). The measure “APIORkBs” refers to the amount of network I/O read traffic in kB per second sent between the servers. The fourth metric, “APIOWkBs,” indicates the amount of network I/O write traffic in kB per second sent between the servers.

Fig. 3 illustrates the average network I/O (KB/s) on each Apache Spark node in tests 1-3 while Fig. 4 demonstrate the average network I/O (KB/s) on each Apache Spark node in tests 3-6.

In the perimeter-based network architecture, the average APRps reads per second are 637 across all Spark worker nodes. The average APWps writes per second are 620. The average APIORkBs read traffic between all Spark worker nodes is 80 while APIOWkBs is 78. The authors reconfigured the network architecture in the subsequent test to provide further insight into IDPS placement impact on distributed big data systems.

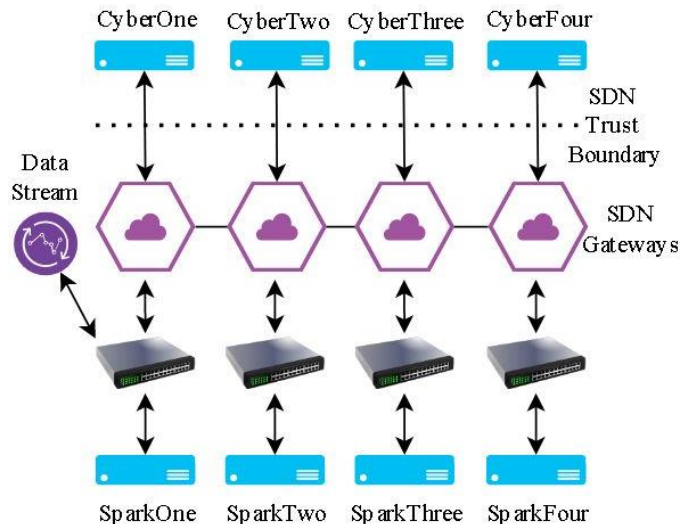


Fig. 2. Perimeter-less security network architecture.

### B. Tests 2-3 Perimeter-less Security Results

Fig. 2 demonstrates the big data network designed for tests two and three. Network architecture uses a modified perimeter-less design proposed by Kotantoulas [39]. In contrast to the traditional perimeter IDPS location in Fig. 1, every big data worker node is in a zero trust network. The authors designed an SD-WAN trust boundary to secure each big data node. The boundary consists of Snort and Suricata intrusion

detection and prevention security gateways. Similar to the virtual software defined perimeter (vEPC) proposed by Bello et al. [40], this study’s zero trust software-based system acts as a security gateway for all distributed servers. Sparkone through Sparkfour are designed to operate securely in most cloud architectures in this model by integrating an SDN security stack on each physically distanced server. The integrated IDPS gateways control and authorize incoming and outgoing network communication. The design emulates the trust boundary surrounding the cloud edge in [39] using the SSH and TLS protocols. Gateways authenticate and connect the distributed systems using a 3072-bit key generated by the Rivest–Shamir–Adleman (RSA) algorithm.

Benchmark results for test 2 with Snort SDN gateways show the wait% is 0.413% and CPU% is 12.54%. Results from this study show that CPU resource consumption is over two times greater in the zero trust architecture than the perimeter network design. Test 3 with Suricata SDN gateways results in 11.05% CPU% and 0.342% wait%. Similar to the perimeter-less design in test 2, test 3 used considerably more CPU resources than test 1. Despite similar rulesets, Suricata SDN gateways used slightly less CPU than Snort.

In the test 2 perimeter-less network architecture the average APRps reads per second are 2,198 across all Spark worker nodes. The average APWps writes per second are 653. The average APIORkBs read traffic between all Spark worker nodes is 298 in test 2, APIOWkBs is 82.

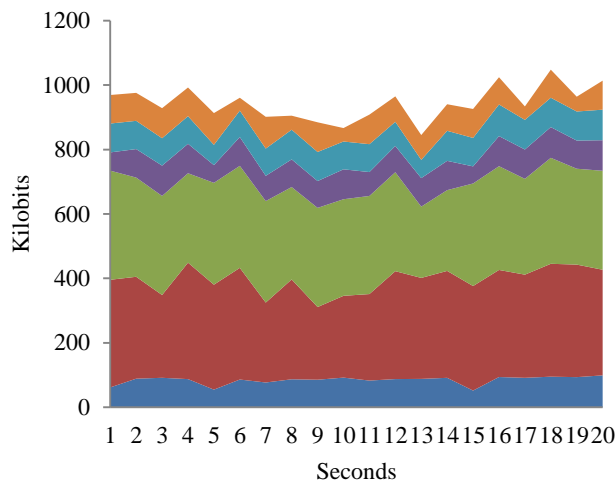


Fig. 3. Tests 1-3 spark per node network I/O in KB/s.

The test 3 network architecture had similar results to test 2. The average APRps reads per second are 2,120 across the distributed Spark systems. The average APWps is 611. APIORkBs between the big data servers is 289 and APIOWkBs is 77. Fig. 3 illustrates the average network I/O (KB/s) on each Apache Spark node in tests 1-3. These results indicate that network traffic and network I/O are nominal when writing to HDFS in all network architectures within this study. In contrast, the number of packets the systems have to read is higher in the perimeter-less network architectures. APRps is over three times higher in tests 2 and 3 than in test 1.

### C. Test 4 Perimeter-Based DDoS Attack Results

Test 4 uses the network architecture (Fig. 1), parallel to test 1. Perimeter-based intrusion detection and prevention systems protect the internal LANs of the Spark nodes. CyberOne through CyberFour are active in test 4. The cyber servers are configured to flood the big data cluster with unlimited TCP SYN handshakes.

Benchmark results for the big data servers during the DDoS attacks parallel test 1 in test 4. In test 4, the IDPSs prevented additional CPU load and network load on the big data servers. In the test case, the hardware IPSs successfully blocked the DDoS attacks.

### D. Tests 5-6 Perimeter-less DDoS Attack Results

Tests 5 and 6 are similar to tests 3 and 4. However, DDoS attacks are administered on the big data cluster. Tests 5-6 use the (Fig. 2) perimeter-less security network architecture. Test 5 uses the Snort-based SDN security boundary, while test 6 uses Suricata. CyberOne through CyberFour are active in tests 5 and 6. The cyber servers execute DDoS attacks on the big data cluster by flooding the servers with unlimited TCP SYN handshakes.

Snort and Suricata security gateways successfully protect the big data systems from DDoS attacks in a zero trust network in tests 5 and 6; however, at the expense of local computational resource increases. Results for test 5 with Snort SDN gateways show the wait% is 0.308% and CPU% is 13.8%. CPU resource consumption increases on average over 1% on the big data servers during the DDoS attacks. Test 6 with Suricata SDN gateways results in 11.95% CPU% and 0.337% wait%. DDoS attacks increased average CPU% by 0.9% across big data systems. Suricata SDN gateways used slightly less CPU than Snort SDN gateways during the DDoS attacks.

Within the test 5 perimeter-less network architecture the average APRps reads per second are 4,762 across all distributed by data secondary nodes. The average APWps writes per second are 626. The average APIORkBs traffic between the distributed systems is 425. APIOWkBs is 79.

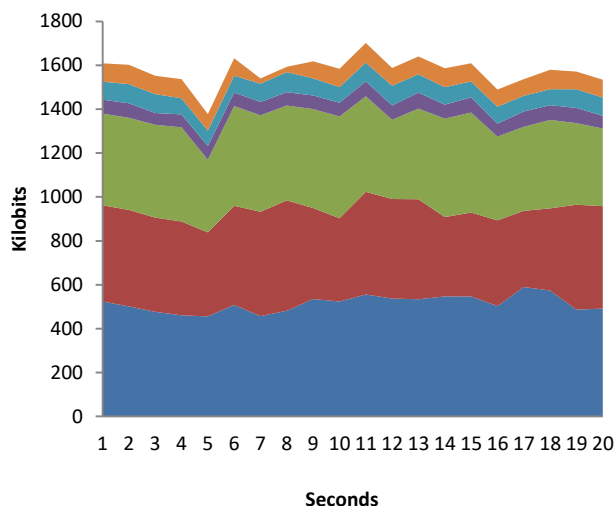


Fig. 4. Tests 4-6 spark per node network I/O in KB/s.

The Suricata gateways in test 6 have average APRps reads per second of 4,311 across the distributed Spark systems. Average APWps is 661. APIORkBs between the big data servers is 416 and APIOWkBs is 81. Fig. 4 demonstrates the average network I/O (KB/s) on each Apache Spark node in tests 3-6.

### E. Test 7 Perimeter-Based DDoS Attack Results

Test 7 shares the same network architecture as test 1 and test 4, illustrated in Fig. 1. To decipher how the DDoS attacks affect the big data servers in the perimeter-based network architecture without IDPS protection, test 7 repeats test 4 but allow all network traffic from CyberOne through CyberFour to the big data cluster. When the DDoS attacks are allowed through the perimeter IPSs in the Fig. 1 network architecture, results show an average CPU% of 17.9% across all distributed big data systems. Predictably, network packets increase in test 7 compared to tests 1 and 4. APRps is 2,895 while APIORkBs is 518. Test 7 has the highest APIORkBs of all network benchmarks performed in this study.

### F. Discussion of the Results

The results illustrate that network traffic and network I/O have marginal differences when writing to HDFS in the network architectures studied. CPU resources and network traffic read by the operating systems increased in zero trust network architectures. The most substantial differences were between tests 4 and 5. During the DDoS attacks, the big data servers required more CPU resources in the perimeter-less security network architecture. In test 5, APIORkBs are considerably higher at 425 than test 4 at 80. This additional traffic is partly due to the SDN security boundaries necessary to protect the systems in a zero trust network environment.

Shifting compute resources closer to individual devices may be necessary as network security perimeters dissipate. However, zero trust architectures in the experimental environment reduced cluster performance. Therefore, additional research is beneficial to optimize the design of perimeter-less network environments.

### G. Limitations

Several environmental factors limit the results. Site-to-site networks were on leased 200 Mbps connections. Future studies might consider leased lines capable of establishing more robust data streams to the distributed nodes. A subsequent restriction is the number of architectures and communication technologies tested. Similar to the architecture in [40], gateways allow for IP Security (IPsec) or Transport Layer Security (TLS) protocols. Future IDPS SDN gateways could add this layer of encryption in a software-defined security boundary between geodistributed big data systems. The outlined limitations emphasize the need for future research to investigate more extensive network architectures and IDPS technologies for big data system security.

## V. CONCLUSION

As the volume of data expands, organizations require big data systems to perform large-scale data analytics. One of several needs for these systems is effective intrusion detection and prevention strategies. This paper builds a review of the

literature on methods used to reduce cybersecurity threats in a range of network architectures that big data systems operate. Findings from literature suggest intrusion detection and prevention systems can respond to certain security attacks. However, a potential disadvantage of capable security systems is the impact on big data system cluster performance. Using a design science approach, the authors develop an eight-step process to benchmark big data systems in varying network architectural environments. The new benchmark process is tested on real-time big data systems running in perimeter-based and perimeter-less network environments. During DDoS cyber-attacks, perimeter-based network architectures outperformed perimeter-less network architectures. This underlines the importance of optimizing the design of zero trust architectures for distributed big data systems.

#### REFERENCES

- [1] D. Gümüşbaş, T. Yıldırım, A. Genovese, and F. Scotti, "A comprehensive survey of databases and deep learning methods for cybersecurity and intrusion detection systems," *IEEE Systems Journal*, vol. 15, no. 2, pp. 1717–1731, Jun. 2021, doi: 10.1109/JSYST.2020.2992966.
- [2] I. D. Aiyanyo, S. Hamman, and H. Lim, "A systematic review of defensive and offensive cybersecurity with machine learning," *Applied Sciences*, vol. 10, no. 17, p. 5811, 2020, doi: 10.3390/app10175811.
- [3] N. V. Patil, C. Rama Krishna, and K. Kumar, "Distributed frameworks for detecting distributed denial of service attacks: A comprehensive review, challenges and future directions," *Concurrency and Computation: Practice and Experience*, vol. 33, no. 10, pp. 1–21, May 2021, doi: 10.1002/cpe.6197.
- [4] R. Rafiq, M. J. Awan, A. Yasin, H. Nobanee, A. M. Zain, and S. A. Bahaj, "Privacy prevention of big data applications: A systematic literature review," *Sage Open*, vol. 12, no. 2, Apr. 2022, doi: 10.1177/21582440221096445.
- [5] R. Mitchell and I. R. Chen, "A survey of intrusion detection techniques for cyber-physical systems," *ACM Comput. Surv.*, vol. 46, no. 4, Mar. 2014, doi: 10.1145/2542049.
- [6] B. B. Zarpelão, R. S. Miani, C. T. Kawakani, and S. C. de Alvarenga, "A survey of intrusion detection in Internet of Things," *Journal of Network and Computer Applications*, vol. 84, pp. 25–37, Apr. 2017, doi: 10.1016/j.jnca.2017.02.009.
- [7] V. Casola, A. De Benedictis, M. Rak, and U. Villano, "Security-by-design in multi-cloud applications: An optimization approach," *Information Sciences*, vol. 454–455, pp. 344–362, Jul. 2018, doi: 10.1016/j.ins.2018.04.081.
- [8] R. Atat, L. Liu, J. Wu, G. Li, C. Ye, and Y. Yang, "Big data meet cyber-physical systems: a panoramic survey," *IEEE Access*, vol. 6, pp. 73603–73636, 2018, doi: 10.1109/ACCESS.2018.2878681.
- [9] R. Gifty, R. Bharathi, and P. Krishnakumar, "Privacy and security of big data in cyber physical systems using Weibull distribution-based intrusion detection," *Neural Computing and Applications*, vol. 31, no. 1, pp. 23–34, Jan. 2019, doi: 10.1007/s00521-018-3635-6.
- [10] S. F. Ochoa, G. Fortino, and G. Di Fatta, "Cyber-physical systems, internet of things and big data," *Future Generation Computer Systems*, vol. 75, pp. 82–84, Oct. 2017, doi: 10.1016/j.future.2017.05.040.
- [11] K. A. P. da Costa, J. P. Papa, C. O. Lisboa, R. Munoz, and V. H. C. de Albuquerque, "Internet of Things: A survey on machine learning-based intrusion detection approaches," *Computer Networks*, vol. 151, pp. 147–157, Mar. 2019, doi: 10.1016/j.comnet.2019.01.023.
- [12] N. Moustafa, B. Turnbull, and K. R. Choo, "An ensemble intrusion detection technique based on proposed statistical flow features for protecting network traffic of Internet of Things," *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4815–4830, Jun. 2019, doi: 10.1109/JIOT.2018.2871719.
- [13] A. Yang, Y. Zhuansun, C. Liu, J. Li, and C. Zhang, "Design of intrusion detection system for Internet of Things based on improved BP neural network," *IEEE Access*, vol. 7, pp. 106043–106052, 2019, doi: 10.1109/ACCESS.2019.2929919.
- [14] Z. Tan *et al.*, "Enhancing big data security with collaborative intrusion detection," *IEEE Cloud Computing*, vol. 1, no. 3, pp. 27–33, Sep. 2014, doi: 10.1109/MCC.2014.53.
- [15] A. N. Jaber and S. U. Rehman, "FCM–SVM based intrusion detection system for cloud computing environment," *Cluster Computing*, vol. 23, no. 4, pp. 3221–3231, Dec. 2020, doi: 10.1007/s10586-020-03082-6.
- [16] M. Hafsa and F. Jemili, "Comparative study between big data analysis techniques in intrusion detection," *Big Data and Cognitive Computing*, vol. 3, no. 1, pp. 1–13, Dec. 2018, doi: 10.3390/bdcc3010001.
- [17] F. M. Awaysheh, M. N. Aladwan, M. Alazab, S. Alawadi, J. C. Cabaleiro, and T. F. Pena, "Security by design for big data frameworks over cloud computing," *IEEE Transactions on Engineering Management*, pp. 1–18, Feb. 2021, doi: 10.1109/TEM.2020.3045661.
- [18] A. Bocci, S. Forti, G. L. Ferrari, and A. Brogi, "Secure FaaS orchestration in the fog: How far are we?" *Computing*, vol. 103, no. 5, pp. 1025–1056, May 2021, doi: 10.1007/s00607-021-00924-y.
- [19] Y. Wang, X. Zhang, Y. Wu, and Y. Shen, "Enhancing leakage prevention for mapreduce," *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 1558–1572, 2022, doi: 10.1109/TIFS.2022.3166641.
- [20] O. Ohrimenko, M. Costa, C. Fournet, C. Gkantsidis, M. Kohlweiss, and D. Sharma, "Observing and preventing leakage in MapReduce," in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, New York, NY, USA, 2015, pp. 1570–1581. doi: 10.1145/2810103.2813695.
- [21] A. M. Sauber, A. Awad, A. F. Shawish, and P. M. El-Kafrawy, "A novel hadoop security model for addressing malicious collusive workers," *Computational Intelligence and Neuroscience*, vol. 2021, pp. 1–10, 2021, doi: 10.1155/2021/5753948.
- [22] P. Derbeko, S. Dolev, E. Gudes, and S. Sharma, "Security and privacy aspects in MapReduce on clouds: A survey," *Computer Science Review*, vol. 20, pp. 1–28, May 2016, doi: 10.1016/j.cosrev.2016.05.001.
- [23] R. Poddar, T. Boelter, and R. Popa, "Arx: An encrypted database using semantically secure encryption," *Proceedings of the VLDB Endowment*, vol. 12, pp. 1664–1678, Jul. 2019, doi: 10.14778/3342263.3342641.
- [24] A. Nisioti, A. Mylonas, P. D. Yoo, and V. Katos, "From intrusion detection to attacker attribution: A comprehensive survey of unsupervised methods," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 4, pp. 3369–3388, Fourthquarter 2018, doi: 10.1109/COMST.2018.2854724.
- [25] S. T. March and G. F. Smith, "Design and natural science research on information technology," *Decision Support Systems*, vol. 15, no. 4, pp. 251–266, Dec. 1995, doi: 10.1016/0167-9236(94)00041-2.
- [26] A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design science in information systems research," *MIS Quarterly*, vol. 28, no. 1, pp. 75–105, 2004, doi: 10.2307/25148625.
- [27] "Dell technology," *Dell Inc*, June, 2022. [Online]. Available: <https://www.dell.com>.
- [28] "Cisco routers and SD-WAN," *Cisco Systems*, June, 2022. [Online]. Available: <https://www.cisco.com/site/us/en/products/networking/sdwan-routers/index.html>.
- [29] "Benchmarking & Diagnostic Software," *Passmark Software*, June, 2022. [Online]. Available: <https://www.passmark.com>.
- [30] "Spark tuning guide on 3rd generation Intel® Xeon® scalable processors based platform," *Intel Corporation*, August, 2021, [Online]. Available: <https://www.intel.cn/content/www/cn/zh/developer/articles/guide/spark-tuning-guide-on-xeon-based-systems.html>.
- [31] "Tuning Spark," *The Apache Software Foundation*, July, 2022. [Online]. Available: <https://spark.apache.org/docs/3.2.2/>.
- [32] "Cluster Mode Overview," *The Apache Software Foundation*, June, 2022. [Online]. Available: <https://spark.apache.org/docs/latest/cluster-overview.html>.
- [33] "Apache Hadoop YARN," *The Apache Software Foundation*, June, 2022. [Online]. Available:

- <https://hadoop.apache.org/docs/stable/hadoop-yarn/hadoop-yarn-site/YARN.html>.
- [34] “Kali linux features,” *OffSec Services Limited*, June, 2022. [Online]. Available: <https://www.kali.org/features>.
- [35] “Snort FAQ/Wiki,” *Cisco Systems*, July, 2022. [Online]. Available: <https://www.snort.org/faq>.
- [36] “Suricata user guide,” *Open Information Security Foundation*, July, 2022. [Online]. Available: <https://suricata.readthedocs.io/en/suricata-6.0.6>.
- [37] “nmon for Linux,” *IBM*, June, 2022. [Online]. Available: <http://nmon.sourceforge.net>.
- [38] “Collapsed data center and campus core deployment guide,” *Cisco Systems*, June, 2022. [Online]. Available: [https://www.cisco.com/c/dam/global/en\\_ca/solutions/strategy/docs/sbaGov\\_nexus7000Dguide\\_new.pdf](https://www.cisco.com/c/dam/global/en_ca/solutions/strategy/docs/sbaGov_nexus7000Dguide_new.pdf).
- [39] J. Kotantoulas, “Zero trust for government networks,” *Cisco Systems*, June, 2022. [Online]. Available: <https://blogs.cisco.com/government/zero-trust-for-government-networks-6-steps-you-need-to-know>.
- [40] Y. Bello, A. R. Hussein, M. Ulema, and J. Koilpillai, “On sustained zero trust conceptualization security for mobile core networks in 5G and beyond,” *IEEE Transactions on Network and Service Management*, vol. 19, no. 2, pp. 1876–1889, Jun. 2022, doi: 10.1109/TNSM.2022.3157248.

# Machine Learning for Smart Cities: A Comprehensive Review of Applications and Opportunities

Xiaoning Dou<sup>1</sup>, Weijing Chen<sup>2\*</sup>, Lei Zhu<sup>3</sup>, Yingmei Bai<sup>4</sup>, Yan Li<sup>5</sup>, Xiaoxiao Wu<sup>6</sup>

School of Business Administration, Xi'an Eurasia University, Xi'an 710065, Shaanxi, China<sup>1,2,4,5</sup>

School of Public Administration, Xi'an University of Architecture and Technology, Xi'an 710054, Shaanxi, China<sup>3</sup>

School of International Education, Southeast Asian University of Thailand, Bangkok 10160, Thailand<sup>6</sup>

**Abstract**—The smart city concept originated a few years ago as a combination of ideas about how information and communication technologies can improve urban life. With the advent of the digital revolution, many cities globally are investing heavily in designing and implementing smart city solutions and projects. Machine Learning (ML) has evolved into a powerful tool within the smart city sector, enabling efficient resource management, improved infrastructure, and enhanced urban services. This paper discusses the diverse ML algorithms and their potential applications in smart cities, including Artificial Intelligence (AI) and Intelligent Transportation Systems (ITS). The key challenges, opportunities, and directions for adopting ML to make cities smarter and more sustainable are outlined.

**Keywords**—Smart city; machine learning; artificial intelligence; intelligent transportation system; smart grids

## I. INTRODUCTION

### A. Background

According to the reports [1], by 2050, the global urban population is expected to reach 70%. This surge in urbanization will drastically impact cities' environment, management, and security. To efficiently handle the meteoric growth in urbanization, many countries have proposed the concept of smart cities to manage resources and optimize energy consumption effectively [2]. Smart city projects can precisely ensure a green environment by developing and adopting low-carbon emission technologies. Urbanization has witnessed unprecedented growth in recent decades, with an increasing number of people migrating to cities for better opportunities and improved quality of life [3]. This rapid urban expansion brings numerous challenges, such as increased energy consumption, traffic congestion, inadequate infrastructure, and environmental degradation [4]. In response, smart cities have emerged as a transformative approach to incorporate advanced Information and Communication Technology (ICT) based hardware and software in urban planning [5]. The smart city utilizes ICT to enhance 'citizens' quality of life, foster the economy, facilitate a process to resolve transport and traffic problems through proper management, encourage a clean and sustainable environment, and provide accessible interaction with the relevant authority of the government [6]. The increased urban expansion and innovations in urban planning and ICT have encouraged planners to focus on promoting the smart city's concept, which considers the well-being of the urban population by focusing on a combination of human, environmental, social, cultural,

energy, information access and usage, and other technological advances [7]. Moreover, as urbanization continues to surge, efficient and sustainable urban public transportation becomes increasingly vital. Association rule mining, a key data analysis technique, plays a crucial role in optimizing public transportation systems by uncovering valuable insights from large datasets. These insights enable cities to enhance transportation efficiency, reduce congestion, and improve overall mobility. The quality of urban public transportation directly impacts the daily lives of millions, affecting everything from commute times to air quality. By harnessing the power of data analytics, cities can provide residents with reliable, accessible, and eco-friendly transportation options, ultimately contributing to improved urban well-being and reduced environmental impact [8].

The proliferation of digital sensors, Internet of Things (IoT) devices, and the availability of vast amounts of data has created new possibilities for harnessing information to optimize urban systems and services [9, 10]. Machine Learning (ML), a branch of Artificial Intelligence (AI), has emerged as a key technology within the smart city context. It enables cities to analyze and extract valuable insights from the vast amounts of data generated by various sources, including sensors, social media, and municipal databases [11]. ML techniques can uncover patterns, correlations, and trends that may go unnoticed, enabling more informed decision-making and proactive interventions [12]. By applying ML algorithms to urban data, cities can gain actionable insights and predictive capabilities in energy management, transportation planning, waste management, public safety, and citizen engagement. These ML-driven applications have the potential to transform traditional urban systems into intelligent, adaptive networks that optimize resource utilization, improve service delivery, and enhance the overall quality of life for residents [13, 14]. However, deploying ML in the complex and dynamic urban environment comes with challenges, ranging from data privacy and security to ensuring ethical and fair AI practices. Addressing these challenges is crucial to realizing the full potential of ML for smart cities and creating sustainable urban ecosystems that meet the evolving needs of residents [15-17].

### B. Literature Review

The emergence of smart cities represents a pivotal response to the challenges posed by rapid urbanization and the increasing demand for improved urban infrastructure and services. As cities grow and evolve, the need to optimize resource management, enhance citizen well-being, and ensure

environmental sustainability has become paramount. This paradigm shift towards smart urbanization is deeply intertwined with the advancements in Information and Communication Technology (ICT) and, more notably, the integration of ML and Artificial Intelligence (AI) into urban planning and governance. In the realm of ML and AI, an extensive body of research has explored their applications across diverse domains, from healthcare to finance and beyond. Within the context of smart cities, these technologies offer unparalleled opportunities for data-driven decision-making, predictive analytics, and automation of urban processes. Studies have demonstrated the potential of ML algorithms in optimizing energy consumption, streamlining transportation systems, enhancing public safety, and promoting sustainable environmental practices. Moreover, ML-driven citizen engagement strategies have shown promise in fostering community collaboration and tailoring services to individual needs [18].

In addition to these technological advancements, the rise of ML in smart cities aligns with broader societal trends, such as the increasing importance of sustainability and the demand for efficient public services. Policymakers, urban planners, and researchers recognize the potential of ML to address the complex and interconnected challenges of modern urban environments [19]. However, the adoption of ML in the urban context is not without its hurdles. Privacy concerns, data quality, ethical considerations, and the need for scalable and interpretable ML models are among the critical issues that warrant careful consideration. This literature review establishes the significance of the research question by highlighting the transformative potential of ML in smart cities, drawing upon existing research and the broader context of urban development. It underscores the need for comprehensive exploration of ML applications, challenges, and opportunities within the smart city framework, which serves as the core focus of this paper.

### C. Objectives

This review paper aims to provide a comprehensive overview of the applications of ML in the context of smart cities. We aim to explore the various ML techniques employed, their impact on urban life, and the challenges and opportunities associated with their implementation. By examining the current state of ML applications in smart cities, we can identify key trends, gaps, and potential future directions for research and development.

### D. Structure of the Review

This paper is organized into several sections to provide a structured analysis of ML applications for smart cities. Section II introduces the foundations of smart cities and ML, highlighting their integration and the potential benefits they offer when combined. Section III explores a range of applications where ML has been successfully applied in smart cities, such as smart energy management, intelligent transportation systems, urban planning and development, public safety and security, waste management and environmental monitoring, healthcare and well-being, and citizen engagement and participation. Section IV highlights future directions and research trends in ML for smart cities,

such as explainable AI, edge computing and distributed ML, federated learning for privacy preservation, IoT integration, and the emergence of urban data marketplaces and governance. Finally, Section V concludes the review by summarizing the key findings, implications, and recommendations for adopting ML in building smarter and more sustainable cities.

## II. FOUNDATIONS OF ML IN SMART CITIES

This section provides a solid foundation by introducing the key concepts and principles underlying smart cities and ML. It offers an overview of the fundamental elements of smart cities, including their objectives, characteristics, and the integration of technology and data-driven approaches. Furthermore, this section explores the core principles and techniques of ML, emphasizing their relevance and applicability within the context of smart cities.

### A. Overview of Smart Cities

Smart cities represent a paradigm shift in urban development, driven by the rapid advancement of technology and the need to address the complex challenges faced by growing urban populations. At its core, a smart city leverages innovative technologies, data analytics, and connectivity to transform urban environments into intelligent, efficient, and sustainable ecosystems. The objectives of smart cities are centered around improving the quality of life for citizens and enhancing the overall efficiency of urban systems [20]. By integrating technology and data, smart cities aim to optimize resource allocation, enhance infrastructure and services, and enable effective decision-making. These objectives are achieved through various domains and initiatives, such as smart governance, smart mobility, smart energy management, smart buildings, and smart healthcare [21, 22].

Smart cities rely on a robust digital infrastructure that supports collecting, storing, and analyzing data from diverse sources [23]. This includes sensors, IoT devices, and communication networks that enable the seamless integration of urban systems. The proliferation of connected devices and the availability of real-time data empower city administrators and residents to make informed decisions and respond quickly to changing circumstances [24]. Moreover, smart cities emphasize citizen-centric approaches, prioritizing the needs and preferences of residents. Through digital platforms and services, citizens can actively participate in decision-making processes, provide feedback and access information about urban services. This promotes community engagement and collaboration, creating more inclusive and responsive urban environments [25]. Smart cities' sustainability is a key pillar as they strive to minimize environmental impact and optimize resource management. This includes initiatives such as smart energy grids, waste management systems, and promoting green and eco-friendly practices. Smart cities aim to reduce carbon emissions, conserve resources, and create a more sustainable future by integrating renewable energy sources, optimizing transportation systems, and implementing efficient waste management strategies [26].

### B. ML Fundamentals

ML is a powerful branch of AI that enables systems to automatically learn from data and make predictions or

decisions without explicit programming. Understanding the fundamentals of ML is essential for comprehending its integration and impact within the context of smart cities [27]. ML algorithms can be categorized into three main types: supervised, unsupervised, and reinforcement learning [28]. In supervised learning, models are trained using labeled data, where the algorithm learns to map input features to corresponding output labels. This approach is commonly used for tasks such as classification and regression [29].

On the other hand, unsupervised learning involves exploring unlabeled data to discover hidden patterns and structures. Clustering and dimensionality reduction are typical applications of unsupervised learning. Reinforcement learning focuses on training an agent to interact with an environment and learn optimal actions based on rewards and feedback. This technique is employed in scenarios where the agent must make sequential decisions [30].

As shown in Fig. 1, the general process of machine learning involves the following key stages. During training, the model learns from a portion of the data by optimizing its internal parameters to minimize prediction errors. The trained model is then evaluated using the testing data to assess its performance and generalization capabilities. Model evaluation metrics, such as accuracy, precision, recall, and F1 score, quantify the model's performance [31]. Feature engineering is a critical aspect of ML, where relevant input features are selected and transformed to improve model performance. This process involves understanding the data, identifying informative features, handling missing values, scaling features, and encoding categorical variables. Ensemble learning techniques, such as bagging and boosting, combine multiple models to make predictions. Transfer learning is another important technique that leverages knowledge gained from one task or domain to improve performance on another related task or domain, reducing the need for extensive training data [32].

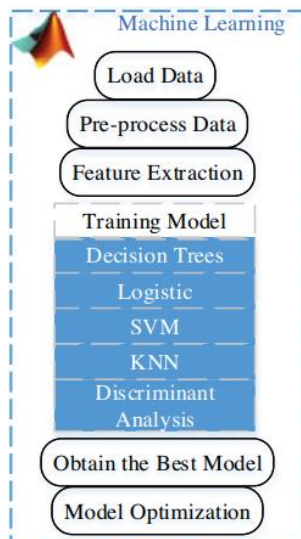


Fig. 1. The general process of ML.

### C. Integration of ML in Smart Cities

Integrating ML techniques within the context of smart cities can revolutionize urban development and enhance

residents' efficiency, sustainability, and quality of life. This subsection explores how ML can be applied in smart cities and the benefits it offers [33]. One key application of ML in smart cities is urban planning and infrastructure management. ML algorithms can analyze vast amounts of data from diverse sources, such as sensor networks, social media, and municipal databases, to gain insights into urban patterns, land use, and transportation flows. These insights enable urban planners to make informed decisions regarding infrastructure development, zoning regulations, and transportation optimization, leading to more efficient and well-designed cities. ML also plays a crucial role in energy management and sustainability within smart cities. By leveraging data from smart grids, energy consumption patterns, and weather forecasts, ML algorithms can optimize energy distribution, predict energy demands, and identify opportunities for energy conservation. This enables cities to reduce energy waste, lower carbon emissions, and promote the integration of renewable energy sources, ultimately contributing to a greener and more sustainable urban environment.

Another important application is in the realm of smart mobility and transportation. ML techniques can analyze real-time data from various sources, including GPS data, traffic cameras, and transportation networks, to predict traffic congestion, optimize route planning, and improve public transportation systems. This leads to reduced congestion, shorter travel times, and enhanced mobility options for citizens. ML also contributes to public safety and security in smart cities. By analyzing data from surveillance systems, social media, and emergency calls, ML algorithms can detect patterns and anomalies, aiding in identifying potential security threats, crime hotspots, and emergency response optimization. This improves the safety and well-being of citizens and enables law enforcement agencies to address security challenges proactively.

Moreover, ML enhances citizen engagement and participation in smart cities. ML algorithms can capture public sentiment, identify community needs, and provide personalized services by analyzing data from social media platforms and citizen feedback. This promotes a sense of inclusion and empowerment among citizens, enabling them to participate in decision-making processes and co-create the urban environment actively. Integrating ML in smart cities brings numerous benefits, including improved urban planning, optimized energy management, enhanced mobility, increased safety, and citizen-centric services. However, challenges such as data privacy, security, ethical considerations, and ensuring fairness in AI algorithms must be addressed to harness the potential of ML in smart cities fully. By overcoming these challenges, cities can leverage the power of ML to create smarter, more sustainable, and livable urban environments.

### III. ML APPLICATIONS IN SMART CITIES

This section presents a clear and comprehensible trend of ML applications in smart cities. As specified in Fig. 2, the potential applications are categorized into seven main categories, including smart city, home automation, and smart healthcare. Tables I to VII summarize the obtained results from reviewing the models.

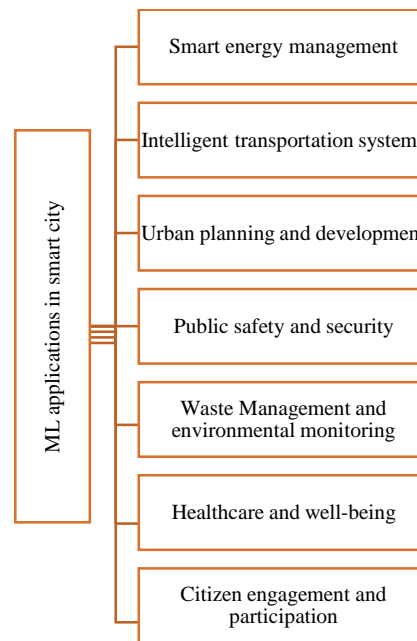


Fig. 2. ML applications in smart cities.

### A. Smart Energy Management

Smart energy management is critical to creating sustainable and efficient smart cities. ML techniques have been successfully applied in various energy management aspects, revolutionizing how energy is generated, distributed, and consumed. In this subsection, we discuss the applications of ML in smart energy management and their impact on creating greener and more efficient urban environments.

- Energy demand prediction: ML algorithms, such as regression models and artificial neural networks, are employed to predict energy demand accurately. By analyzing historical energy consumption data, weather patterns, and other relevant factors, these models can accurately forecast future energy demand. This information enables utility providers to optimize energy production and distribution, ensuring a reliable and efficient energy supply while minimizing waste.
- Energy load forecasting: ML techniques are used to forecast energy load patterns in real-time. By analyzing data from smart meters, weather conditions, and historical load profiles, algorithms can predict future load patterns. This information aids in managing peak demand, optimizing energy distribution, and facilitating the integrating of renewable energy sources into the grid. Load forecasting helps utilities balance supply and demand, reduce costs, and improve the overall reliability and stability of the energy grid.
- Energy optimization and control: ML algorithms optimize energy consumption within smart buildings and homes. By leveraging data from sensors, occupancy

patterns, and weather conditions, algorithms can learn and adapt to energy usage patterns. They can automatically adjust heating, cooling, and lighting systems to optimize energy efficiency while maintaining occupant comfort. Energy optimization algorithms help reduce energy waste, lower utility bills, and promote sustainable energy consumption practices.

- Energy theft detection: The ML techniques aid in detecting energy theft and unauthorized usage within the energy grid. By analyzing consumption patterns and identifying anomalies, algorithms can flag suspicious activities that indicate potential theft or tampering. This helps utility companies prevent revenue loss and ensure fair distribution of energy resources.
- Renewable energy integration: ML is crucial in integrating renewable energy sources into the energy grid. Algorithms can analyze weather data, historical renewable energy generation, and demand patterns to optimize the utilization and management of renewable energy resources. This enables effective grid integration, reduces reliance on fossil fuels, and promotes the transition to a greener, more sustainable energy infrastructure.

ML applications in smart energy management offer significant benefits such as improved energy efficiency, cost savings, reduced carbon emissions, and enhanced grid reliability. However, data quality, privacy, and algorithmic transparency challenges need to be addressed to ensure the responsible and effective deployment of ML techniques in smart cities' energy systems.



TABLE I. ML APPLICATIONS IN SMART ENERGY MANAGEMENT

Approach	ML type	Objective	Achievement	Challenges	References
Energy demand prediction	Supervised	Predict future energy demand accurately using ML algorithms	Accurate forecasting enables the optimization of energy distribution, cost reduction, and efficient load balancing	Relies on historical data and assumptions and may not account for sudden changes or events that deviate from historical patterns Requires continuous updating and validation to account for evolving energy consumption patterns	[34-40]
Energy load forecasting	Supervised	Forecast real-time energy load patterns based on data from smart meters and weather conditions	Enables efficient energy distribution, demand management, and integration of renewable energy sources Helps balance supply and demand, improve grid stability, and optimize resource allocation	Relies on accurate and timely data from smart meters and weather sensors Uncertainty in weather conditions and unforeseen events can impact the accuracy of load forecasts Requires robust data management and monitoring systems to ensure data quality and reliability	[41-48]
Energy optimization and control	Supervised	Optimize energy consumption in smart buildings and homes through ML algorithms.	Maximizes energy efficiency, reduces waste, and lowers utility bills Improves occupant comfort by dynamically adjusting heating, cooling, and lighting systems Enables demand response strategies and load balancing	Requires integration with smart devices and sensors for real-time data collection Dependency on accurate data and system feedback Potential privacy concerns related to the collection and usage of personal data Optimization algorithms may face challenges in highly dynamic environments and require continuous adaptation to changing conditions.	[49-53]
Energy theft detection	Supervised	Detect and flag potential energy theft or unauthorized usage within the energy grid using ML.	Helps prevent revenue loss and ensure fair energy distribution Improves the financial sustainability of utility providers Identifies anomalies and patterns indicative of energy theft or tampering	Relies on data quality and availability. False positives or false negatives may occur, requiring human intervention for verification May face challenges in identifying sophisticated or evolving techniques used for energy theft	[54-60]
Renewable energy integration	Supervised	Optimize the integration of renewable energy sources into the energy grid through ML algorithms.	Enables efficient utilization of renewable energy, reduces reliance on fossil fuels, and lowers carbon emissions Optimizes resource allocation based on weather patterns, demand, and grid conditions	Relies on accurate weather data and renewable energy generation forecasts. Uncertainty in weather patterns The intermittent nature of renewable sources can pose challenges in balancing supply and demand. Integrating diverse renewable sources and their variability may require advanced modeling and management strategies.	[61-64]

### B. Intelligent Transportation Systems

Intelligent Transportation System (ITS) plays a crucial role in enhancing urban transportation's efficiency, safety, and sustainability. ML techniques have been widely applied in various aspects of ITS to optimize traffic management, improve transportation infrastructure, and provide intelligent decision-making capabilities. In this subsection, we discuss the applications of ML in ITS and their impact on creating smarter and more efficient urban mobility.

- Traffic prediction and management: ML algorithms predict and manage traffic flow in real-time. These algorithms can forecast traffic patterns and congestion levels by analyzing historical traffic data, weather, and other relevant factors. This information aids in proactive traffic management, optimizing signal

timings, rerouting strategies, and providing real-time traffic updates to drivers and traffic management authorities. ML-based traffic prediction and management systems improve traffic flow, reduce congestion, and enhance overall transportation efficiency.

- Intelligent routing and navigation: ML techniques enable intelligent routing and navigation systems considering real-time traffic conditions, road incidents, and user preferences. These systems use ML algorithms to analyze historical and real-time data, such as traffic flow, accidents, and road closures, to provide optimal routes to drivers. By considering dynamic factors, ML-based routing and navigation systems help reduce travel time, fuel consumption, and environmental impact, improving overall transportation efficiency.

- **Vehicle and pedestrian safety:** ML algorithms contribute to improving vehicle and pedestrian safety in smart cities. Combined with ML, computer vision techniques enable intelligent video surveillance systems to detect and analyze traffic violations, identify potential safety risks, and provide early warning alerts. ML algorithms can also analyze vehicle sensor data to predict and prevent accidents by detecting anomalies, identifying aggressive driving behavior, and supporting advanced driver assistance systems (ADAS). These applications enhance road safety, reduce accidents, and improve transportation security.
- **Public transportation optimization:** ML techniques optimize public transportation systems, including bus and train schedules, route planning, and fleet management. ML algorithms can optimize public transportation services, improve reliability, reduce waiting times, and enhance passenger satisfaction by analyzing historical ridership data, weather conditions, and other factors. ML algorithms can also support demand-responsive transportation systems, enabling adaptive routing and scheduling based on real-time demand and passenger preferences.
- **Smart parking management:** ML algorithms are used to optimize parking management in smart cities. By analyzing data from sensors, historical occupancy patterns, and real-time information, ML-based parking

systems can provide accurate parking availability predictions, guide drivers to available parking spaces, and optimize parking space utilization. These applications reduce traffic congestion, lower vehicle emissions, and improve the overall efficiency of parking operations.

- ML applications in ITS offer significant benefits, including improved traffic flow, enhanced transportation efficiency, increased safety, and reduced environmental impact. However, data privacy, scalability, and algorithmic transparency must be addressed to ensure the responsible and effective deployment of ML techniques in smart city transportation systems. Ongoing research and development efforts aim to overcome these challenges and unlock the full potential of ML in shaping the future of urban mobility.

### C. Urban Planning and Development

Urban planning and development play a vital role in shaping cities' physical and social infrastructure. ML techniques have emerged as powerful tools for analyzing vast data and extracting valuable insights to support urban planning and development decisions [86]. In this subsection, we discuss the applications of ML in smart cities' urban planning and development and how they contribute to creating sustainable, livable, and efficient urban environments.

TABLE II. ML APPLICATIONS IN ITS

Approach	ML type	Objective	Achievement	Challenges	References
Traffic prediction and management	Supervised	ML algorithms predict and manage traffic flow in real-time, optimizing signal timings and providing updates.	Improved traffic flow Reduced congestion Proactive management.	Relies on accurate and up-to-date data, challenges in data integration and availability Limited control over external factors like accidents or road works	[65-71]
Intelligent routing and navigation	Supervised	ML enables intelligent routing systems to consider real-time traffic conditions, incidents, and user preferences.	Reduced travel time, fuel consumption, and environmental impact Improved navigation and route optimization	Dependency on accurate and real-time data, challenges in integrating multiple data sources Potential biases in data can lead to suboptimal route recommendations	[72-75]
Vehicle and pedestrian safety	Supervised	ML-based surveillance systems detect traffic violations, identify risks, and support driver assistance systems	Improved road safety Early warning alerts Accident prevention	Challenges in real-time detection accuracy Potential privacy concerns related to surveillance systems limitations in detecting complex traffic scenarios or unpredictable pedestrian behavior	[76-78]
Public transportation optimization	Supervised	ML optimizes public transportation systems, schedules, and route planning based on demand and historical data.	Enhanced public transportation services Improved reliability and passenger satisfaction	Data integration challenges Limited effectiveness during unexpected events or disruptions, dependency on accurate and up-to-date ridership data	[79-81]
Smart parking management	Supervised	ML algorithms optimize parking space utilization and guide drivers to available parking spaces.	Reduced traffic congestion Improved parking efficiency Lower vehicle emissions.	Dependence on accurate and real-time parking occupancy data Challenges in sensor deployment and maintenance Limited effectiveness in highly congested areas.	[82-85]

- Land use and zoning optimization: ML algorithms analyze various data sources, such as satellite imagery, demographic data, and economic indicators, to optimize land use and zoning regulations. By identifying patterns and relationships in data, ML can assist urban planners in determining the most suitable locations for residential, commercial, and industrial zones. These insights enable more efficient land use planning, balanced development, and the promotion of mixed-use neighborhoods.
  - Transportation infrastructure planning: ML techniques aid in transportation infrastructure planning by analyzing data on population distribution, commuting patterns, and transportation demand. These algorithms can identify optimal locations for transportation hubs, such as bus stops, metro stations, or bike-sharing stations, based on demand and accessibility factors. ML-based transportation planning improves connectivity, reduces travel time, and enhances transportation efficiency.
  - Environmental impact assessment: ML algorithms are employed to assess the environmental impact of urban development projects. These algorithms can predict the potential impact of proposed projects by analyzing air quality, noise levels, water resources, and biodiversity. This information assists in making informed decisions, ensuring sustainable development practices, and minimizing negative environmental effects.
  - Urban mobility and traffic management: ML techniques optimize urban mobility and traffic management by analyzing data from various sources, including sensors, GPS devices, and social media feeds. These algorithms can identify traffic patterns, predict congestion, and optimize transportation routes and signals. ML-based traffic management systems enhance traffic flow, reduce congestion, and improve the overall efficiency of urban transportation.
  - Infrastructure maintenance and management: ML algorithms contribute to maintaining and managing urban infrastructure, such as roads, bridges, and utilities. These algorithms analyze sensor data, maintenance records, and historical patterns to predict infrastructure deterioration and schedule maintenance activities. ML-based systems help ensure urban infrastructure reliability, safety, and longevity by optimizing maintenance efforts.
- ML applications in urban planning and development provide significant benefits, including optimized land use, improved transportation infrastructure, sustainable development practices, and efficient management of urban assets. However, challenges such as data quality, data integration, and interpretability of ML models must be addressed to ensure the effective and responsible application of ML techniques in urban planning processes. Ongoing research and collaboration between urban planners and data scientists aim to overcome these challenges and leverage the full potential of ML in shaping smarter and more sustainable cities.

TABLE III. ML APPLICATIONS IN URBAN PLANNING AND DEVELOPMENT

Approach	ML type	Objective	Achievement	Challenges	References
Land use and zoning optimization	Supervised	Data is analyzed to optimize land use and zoning regulations for sustainable development.	Efficient land use planning Promotion of mixed-use neighborhoods Optimized resource allocation	Relies on accurate and comprehensive data Challenges in integrating various data sources Potential biases in data affecting zoning decisions	[87, 88]
Transportation infrastructure planning	Supervised	Transportation infrastructure is optimized by identifying optimal locations for hubs and facilities.	Enhanced connectivity Reduced travel time Improved transportation efficiency	Dependency on accurate and up-to-date data Challenges in integrating different transportation modes Potential biases in data affecting planning decisions	[89, 90]
Environmental impact assessment	Supervised	The environmental impact of development projects is assessed based on various data sources.	Informed decision-making Promotion of sustainable development practices Reduced environmental impact	Relies on accurate and comprehensive environmental data Challenges in quantifying long-term environmental impacts Potential biases in data affecting assessments	[91-93]
Urban mobility and traffic management	Supervised	Urban mobility and traffic management are optimized by analyzing data from various sources.	Improved traffic flow Reduced congestion Enhanced transportation efficiency	Dependency on accurate and real-time data Challenges in data integration and processing Potential biases in data affecting traffic management decisions	[94-96]
Infrastructure maintenance and management	Supervised	Infrastructure deterioration and maintenance activities are predicted	Enhanced infrastructure reliability Optimized maintenance scheduling Improved asset management	Relies on accurate infrastructure data Challenges in integrating maintenance records Potential biases in data affecting maintenance decisions	[97-99]

#### D. Public Safety and Security

Ensuring public safety and security is a critical aspect of smart city initiatives. ML techniques have emerged as powerful tools in analyzing large volumes of data and extracting meaningful insights to enhance public safety measures and security systems. In this subsection, we discuss the applications of ML in smart cities' public safety and security domains and how they contribute to creating safer and more secure urban environments.

- **Video surveillance and monitoring:** ML algorithms enable intelligent video surveillance systems that can analyze real-time video feeds from cameras across the city. These algorithms can automatically detect and track suspicious activities, identify objects of interest, and raise alerts for potential security threats. ML-based video surveillance enhances situational awareness, improves incident response, and aids in crime prevention and detection.
- **Predictive policing:** ML techniques are employed to predict and prevent crime by analyzing historical crime data, socio-economic indicators, and other relevant factors. These algorithms can identify patterns, hotspots, and trends, enabling law enforcement agencies to deploy resources strategically and proactively. ML-based predictive policing helps reduce crime rates, improve resource allocation, and enhance public safety.
- **Emergency response optimization:** ML algorithms optimize emergency response systems by analyzing emergency call records, traffic conditions, and geographical information. These algorithms can identify the optimal deployment of emergency vehicles, predict response times, and dynamically allocate resources based on real-time incidents. ML-based emergency response systems improve response efficiency, minimize response times, and save lives in critical situations.
- **Cybersecurity and threat detection:** ML techniques aid in cybersecurity and threat detection by analyzing network traffic, user behavior, and system logs to detect anomalies and potential security breaches. These algorithms can identify patterns of malicious activity, classify threats, and provide early warnings to prevent cyber-attacks. ML-based cybersecurity systems protect critical infrastructure, sensitive data, and digital services.
- **Disaster management and resilience:** ML algorithms contribute to disaster management and resilience by analyzing data from various sources, such as weather forecasts, sensor networks, and social media feeds. These algorithms can predict and model the impact of natural disasters, aid in evacuation planning, and assist in resource allocation during emergencies. ML-based disaster management systems enhance preparedness, response, and recovery capabilities.

TABLE IV. ML APPLICATIONS IN PUBLIC SAFETY AND SECURITY

Approach	ML type	Objective	Achievement	Challenges	References
Video surveillance and monitoring	Supervised	Real-time video feeds are analyzed to detect and track suspicious activities and objects.	Enhanced situational awareness Improved incident response Crime prevention and detection	Dependency on accurate and high-quality video feeds Potential biases in the algorithmic analysis Privacy concerns related to extensive video surveillance	[100-102]
Predictive policing	Supervised	Crime is predicted and prevented by analyzing historical data and relevant socio-economic factors.	Proactive resource allocation Reduced crime rates Improved law enforcement strategies	Relies on accurate and comprehensive data Potential biases in data affecting predictions Ethical concerns related to algorithmic profiling	[103, 104]
Emergency response optimization	Supervised	Emergency response systems are optimized by predicting response times and resource allocation.	Efficient resource allocation Reduced response times Improved emergency management	Dependency on accurate and real-time data Challenges integrating multiple data sources Potential biases in algorithmic predictions	[105-107]
Cybersecurity and threat detection	Supervised	Network traffic and user behavior are analyzed to detect and prevent cyber threats and breaches.	Early detection of anomalies Improved threat prevention Enhanced critical infrastructure protection	Evolving nature of cyber threats Challenges in identifying new and sophisticated attack patterns Potential biases in algorithmic analysis	[108-110]
Disaster management and resilience	Supervised	Data is analyzed to predict and manage the impact of natural disasters and aid in recovery efforts.	Improved preparedness and response Enhanced resource allocation Efficient evacuation planning	Dependency on accurate and comprehensive data Challenges integrating various data sources Potential biases in algorithmic predictions related to complex disaster scenarios	[111-113]

ML applications in smart cities' public safety and security domains offer significant benefits, including improved situational awareness, proactive crime prevention, efficient emergency response, enhanced cybersecurity, and better disaster management. However, challenges such as data privacy, algorithmic biases, and ethical considerations need to be addressed to ensure the responsible and effective deployment of ML techniques in public safety and security systems. Ongoing research and collaboration between law enforcement agencies, security experts, and data scientists aim to overcome these challenges and leverage the full potential of ML in creating safer and more secure smart cities.

*E. Waste Management and Environmental Monitoring*

Effective waste management and environmental monitoring are essential for smart city initiatives to create sustainable, eco-friendly urban environments. ML techniques have revolutionized these areas by enabling the analysis of large-scale data sets and extracting valuable insights for optimizing waste management processes and monitoring environmental conditions. In this subsection, we discuss the applications of ML in smart cities' waste management and environmental monitoring and how they contribute to achieving efficient resource utilization and environmental sustainability.

- **Waste sorting and recycling:** ML algorithms play a crucial role in waste sorting and recycling by automating the identification and segregation of different waste materials. Using computer vision and image recognition techniques, these algorithms can analyze images of waste and classify them into specific categories, such as plastic, paper, glass, or organic

waste. ML-based waste sorting systems enhance recycling efforts, reduce landfill waste, and promote a circular economy.

- **Predictive waste collection:** ML techniques optimize waste collection routes and schedules based on predictive analysis. By analyzing historical data on waste generation patterns, population density, and other relevant factors, these algorithms can predict the optimal time and location for waste collection. ML-based waste collection systems reduce operational costs, minimize environmental impact, and improve efficiency.
- **Environmental quality monitoring:** ML algorithms analyze data from environmental sensors and monitoring devices to assess air quality, water quality, noise levels, and other environmental parameters. These algorithms can detect patterns, identify pollution sources, and predict environmental risks. ML-based environmental monitoring systems facilitate early detection of pollution events, enable targeted interventions, and promote healthier and cleaner urban environments.
- **Energy optimization and conservation:** ML techniques optimize energy consumption and promote energy conservation in smart cities. These algorithms analyze data on energy usage patterns, weather conditions, and building characteristics to identify opportunities for energy savings. ML-based energy management systems can dynamically adjust energy usage, optimize building operations, and promote sustainable energy practices.

TABLE V. ML APPLICATIONS IN WASTE MANAGEMENT AND ENVIRONMENTAL MONITORING

Approach	ML type	Objective	Achievement	Challenges	References
Waste sorting and recycling	Supervised	The identification and sorting of waste materials are automated for recycling.	Improved recycling efforts Reduced landfill waste Promoting a circular economy	Dependency on accurate and comprehensive waste data Challenges in integrating waste sorting systems Potential biases in algorithmic classification	[114-117]
Predictive waste collection	Supervised	Waste collection routes and schedules are optimized based on predictive analysis.	Reduced operational costs Minimized environmental impact Improved waste management efficiency	Dependency on accurate waste generation data Challenges in integrating real-time data Potential biases in algorithmic predictions	[118]
Environmental quality monitoring	Supervised	Data is analyzed from environmental sensors to assess air quality, water quality, and noise levels.	Early detection of pollution events Targeted interventions Promotion of healthier urban environments	Relies on accurate and comprehensive environmental data, sensor deployment, maintenance challenges Potential biases in algorithmic analysis.	[119]
Energy optimization and conservation	Supervised	Energy consumption is optimized, and energy conservation is promoted in smart cities.	Reduced energy usage Improved energy management Promoted sustainable energy practices	Dependency on accurate and real-time energy data Challenges in integrating heterogeneous data sources Potential biases in algorithmic optimization	[120, 121]
Green spaces management	Supervised	ML algorithms optimize the management of green spaces by analyzing data on soil moisture and plant health.	Efficient resource management, water conservation, and promotion of healthy urban ecosystems	Relies on accurate and comprehensive data on soil and plant conditions, data collection, and maintenance challenges Potential biases in algorithmic analysis.	[122]

- Green spaces management: ML algorithms contribute to efficiently managing green spaces, such as parks and gardens, by analyzing data on soil moisture, weather conditions, and plant health. These algorithms can optimize irrigation schedules, detect plant disease outbreaks, and support precision agriculture techniques. ML-based green space management systems enhance resource efficiency, conserve water, and promote healthy urban ecosystems.

ML applications in waste management and environmental monitoring offer significant benefits, including improved waste sorting and recycling, optimized waste collection processes, enhanced environmental quality monitoring, energy conservation, and efficient management of green spaces. However, challenges such as data quality, integration of heterogeneous data sources, and interpretability of ML models need to be addressed to ensure the effective and responsible deployment of ML techniques in these domains. Ongoing research and collaboration between waste management experts, environmental scientists, and data scientists aim to overcome these challenges and leverage the full potential of ML in creating sustainable and environmentally conscious smart cities.

#### F. Healthcare and Well-being

The application of ML in the healthcare and well-being domain of smart cities has the potential to revolutionize the delivery of healthcare services, improve patient outcomes, and enhance overall well-being. ML techniques enable the analysis of large volumes of healthcare data, including patient records, medical images, and sensor data, to extract valuable insights and support personalized and proactive healthcare interventions. In this subsection, we discuss the applications of ML in smart cities' healthcare and well-being domains and how they contribute to creating healthier and more resilient urban communities.

- Disease diagnosis and predictive analytics: ML algorithms can analyze patient data, such as symptoms, medical history, and test results, to aid disease diagnosis and prediction. These algorithms can identify patterns, detect anomalies, and provide early disease warnings, enabling timely interventions and personalized treatment plans. ML-based diagnostic systems improve accuracy, reduce misdiagnosis, and enhance patient care.
- Remote patient monitoring: ML techniques enable remote monitoring of patients' health conditions using wearable devices and sensors. These algorithms can analyze real-time data, such as heart rate, blood pressure, and activity levels, to detect deviations from normal patterns and alert healthcare providers. ML-based remote monitoring systems facilitate proactive interventions, reduce hospitalizations, and enhance patient convenience and comfort.
- Health risk assessment and prevention: ML algorithms analyze various data sources, including lifestyle data, environmental factors, and genetic information, to assess individuals' health risks and provide personalized

recommendations for prevention. These algorithms can identify risk factors, predict susceptibility to diseases, and suggest healthy lifestyle interventions. ML-based health risk assessment systems empower individuals to make informed decisions, promote preventive care, and reduce healthcare costs.

- Health resource optimization: ML techniques optimize the allocation of healthcare resources, such as hospital beds, medical staff, and equipment. These algorithms can analyze patient data, bed occupancy rates, and historical trends to predict future demand and facilitate resource planning. ML-based resource optimization systems improve operational efficiency, reduce waiting times, and ensure better utilization of healthcare resources.
- Mental health support: ML algorithms contribute to mental health support by analyzing data from various sources, such as social media posts, wearable devices, and electronic health records. These algorithms can detect patterns indicative of mental health conditions, provide personalized recommendations, and offer virtual counseling and support. ML-based mental health support systems enhance access to care, reduce stigma, and improve mental well-being in smart cities.

The applications of ML in the healthcare and well-being domains of smart cities offer significant benefits, including improved disease diagnosis, proactive healthcare interventions, personalized treatment plans, optimized resource allocation, and enhanced mental health support. However, challenges such as data privacy and security, ethical considerations, and biases in algorithmic analysis need to be addressed to ensure the responsible and effective deployment of ML techniques in these domains. Ongoing research and collaboration between healthcare professionals, data scientists, and policymakers aim to overcome these challenges and harness the full potential of ML in creating healthier and more resilient smart cities.

#### G. Citizen Engagement and Participation

Citizen engagement and participation are key components of smart cities, aiming to involve residents in decision-making processes and improve the quality of urban life. ML techniques significantly facilitate citizen engagement by analyzing large amounts of data and enabling personalized interactions between citizens and city authorities. In this subsection, we discuss the applications of ML in smart cities' citizen engagement and participation domains, highlighting how they enhance residents' communication, collaboration, and empowerment.

- Sentiment analysis and feedback processing: ML algorithms analyze public sentiment by mining social media posts, online reviews, and citizen feedback. These algorithms can automatically classify sentiments as positive, negative, or neutral, providing valuable insights into public opinions about various aspects of urban life. Sentiment analysis helps city authorities understand citizen concerns, identify areas for improvement, and tailor policies and services accordingly.

TABLE VI. ML APPLICATIONS IN HEALTHCARE AND WELL-BEING

Approach	ML type	Objective	Achievement	Challenges	References
Disease diagnosis and predictive analytics	Supervised	Patients' data are analyzed to aid in disease diagnosis and prediction	Improved accuracy Early detection of diseases Personalized treatment plans	Dependency on accurate and comprehensive patient data Potential biases in algorithmic analysis Challenges in interpretability	[123-126]
Remote patient monitoring	Supervised	Remote monitoring of patients' health conditions using wearable devices and sensors	Proactive interventions Reduced hospitalizations Improved convenience for patients	Reliability and accuracy of sensor data Potential privacy concerns Challenges in data integration.	[127-129]
Health risk assessment and prevention	Supervised	Individuals' health risks are assessed, and personalized recommendations are provided for prevention.	Personalized recommendations for preventive care, Reduced healthcare costs	Reliance on accurate and diverse data sources Potential biases in algorithmic analysis Ethical considerations	[130-132]
Health resource optimization	Supervised	The allocation of healthcare resources is optimized based on patient data and demand predictions.	Improved resource utilization Reduced waiting times Efficient resource planning	Data accuracy and quality Challenges in integrating multiple data sources Potential biases in demand predictions	[133, 134]
Mental health support	Supervised	Various data sources are analyzed to provide mental health support and virtual counseling.	Enhanced access to care Reduced stigma Personalized support for mental well-being	Privacy and security concerns related to sensitive mental health data Potential biases in algorithmic analysis	[135, 136]

- **Participatory decision-making:** ML techniques enable participatory decision-making by providing platforms for citizens to express their opinions, vote on proposals, and contribute to policy development. These algorithms can aggregate and analyze citizen inputs, allowing city authorities to make informed decisions that reflect the preferences and priorities of the community. Participatory decision-making enhances transparency, accountability, and democratic processes in smart cities.
- **Personalized citizen services:** ML algorithms personalize citizen services by leveraging data on individual preferences, behaviors, and needs. These algorithms can recommend relevant information, services, and events based on citizens' profiles and historical interactions. Personalization enhances citizen experience, increases engagement, and fosters a sense of belonging in the community.
- **Urban analytics and planning:** ML techniques analyze data from various sources, including sensors, traffic patterns, and citizen-generated data, to generate urban planning and development insights. These algorithms can identify usage patterns, predict future trends, and optimize urban infrastructure and services. Urban analytics and planning empower city authorities to make data-driven decisions, improve resource

allocation, and create more livable and sustainable cities.

- **Community empowerment and collaboration:** ML algorithms facilitate community empowerment and collaboration by connecting citizens with similar interests and promoting collective action. These algorithms can identify common goals, facilitate collaboration platforms, and support grassroots initiatives. Community empowerment enhances social cohesion, fosters civic engagement, and encourages residents to actively participate in shaping their neighborhoods.

ML applications in citizen engagement and participation domains of smart cities offer significant benefits, including improved communication between citizens and city authorities, participatory decision-making, personalized citizen services, data-driven urban planning, and community empowerment. However, challenges such as data privacy, the digital divide, biases in algorithmic analysis, and ensuring inclusive participation need to be addressed to ensure equitable and meaningful engagement of all residents. Ongoing research and collaboration between data scientists, urban planners, and policymakers aim to overcome these challenges and leverage the full potential of ML in enhancing citizen engagement and building inclusive smart cities.

TABLE VII. ML APPLICATIONS IN CITIZEN ENGAGEMENT AND PARTICIPATION

Approach	ML type	Objective	Achievement	Challenges	References
Sentiment analysis and feedback processing		Public sentiment and citizen feedback are analyzed to gain insights into public opinions and concerns.	Understand citizen sentiments: Sentiment analysis allows for the comprehension of public sentiments and concerns, aiding policymakers in making informed decisions Identify areas for improvement: By processing citizen feedback, areas for policy improvement can be pinpointed, leading to more effective governance Tailor policies accordingly: Tailoring policies to address specific citizen sentiments enhances public satisfaction and engagement.	Biases in sentiment analysis: Ensuring the accuracy and impartiality of sentiment analysis remains a challenge Challenges in handling unstructured data: Managing and extracting insights from unstructured data, such as text and social media content, require advanced techniques Potential privacy concerns related to data mining: Ethical considerations surrounding data mining must be addressed to protect citizen privacy.	[137-140]
Participatory decision-making		Enable citizens to participate in decision-making processes and contribute to policy development actively.	Increased transparency, accountability, and democratic processes: Involving citizens in decision-making enhances government transparency and accountability Representation of citizen preferences and priorities: Decision-making reflects the diverse preferences and priorities of the community, leading to more inclusive policies.	Digital divide: Ensuring equitable access to participation platforms and overcoming the digital divide is essential for meaningful engagement Potential biases in algorithmic analysis, ensuring inclusivity and diversity in participation: Care must be taken to mitigate algorithmic biases and encourage diverse citizen participation.	[141, 142]
Personalized citizen services		Citizen services are personalized by recommending relevant information, events, and services based on profiles.	Enhanced citizen experience: Personalization improves the user experience and increases citizen engagement Increased engagement: Tailored recommendations encourage citizens to interact more with available services Tailored services: Citizens receive services that match their specific needs and interests.	Privacy concerns related to data collection and profiling: Safeguarding citizen privacy in data collection and profiling processes is critical Potential biases in personalization algorithms: Ensuring that personalization algorithms do not reinforce biases is an ongoing challenge.	[143, 144]
Urban analytics and planning		Urban data is analyzed to generate insights for urban planning, infrastructure optimization, and resource allocation.	Data-driven decision-making: Urban analytics facilitates data-driven decision-making, leading to more efficient resource allocation and planning. Optimized resource allocation: Through data analysis, cities can allocate resources more effectively, reducing waste Improved urban infrastructure and services: Data-driven insights enhance the quality of urban services and infrastructure.	Data quality and integration: Ensuring the accuracy and integration of data from various sources is vital for meaningful analysis Interpretability of ML models: Understanding how ML models arrive at conclusions is crucial for decision-makers Biases in data and algorithms: Identifying and addressing biases in data and algorithms is essential to avoid unintended consequences.	[145-147]
Community empowerment and collaboration		Community empowerment and collaboration by connecting citizens and supporting collective action	Foster social cohesion: Connecting citizens fosters social cohesion and a sense of community Encourage civic engagement: Empowering citizens to take action encourages active civic participation. Support grassroots initiatives: ML algorithms can connect citizens with grassroots initiatives that align with their interests.	Ensuring inclusive participation: Efforts must be made to ensure that all segments of the population have opportunities to engage. Potential biases in algorithmic matchmaking: Algorithms must be designed to avoid excluding certain groups inadvertently. Challenges sustaining community engagement and collaboration: Sustaining long-term community engagement requires ongoing effort and commitment.	[148]

#### IV. FUTURE DIRECTIONS AND RESEARCH TRENDS

ML applications in smart cities are constantly evolving, and several future directions and research trends hold promise for advancing the capabilities and impact of smart city technologies. In this subsection, we discuss some key areas likely to shape the future of ML in smart cities.

- Explainability and transparency: As ML algorithms become more complex and pervasive in smart cities, there is a growing need for explainability and transparency. Researchers are exploring techniques to make ML models more interpretable, allowing stakeholders to understand the reasoning behind algorithmic decisions. Ensuring transparency not only



builds trust but also helps in identifying potential biases and addressing ethical concerns.

- **Privacy and security:** With the increasing use of data in smart city environments, preserving privacy and ensuring data security are critical research areas. Future work aims to develop robust privacy-preserving ML techniques for data analysis while protecting sensitive information. Additionally, efforts are focused on enhancing the security of ML models to prevent adversarial attacks and unauthorized access to data.
- **Federated learning and edge computing:** Federated learning, a distributed learning approach, is gaining attention in the context of smart cities. It allows training ML models on decentralized data sources while preserving data privacy. Furthermore, integrating ML with edge computing enables real-time data processing and decision-making at the network's edge, reducing latency and dependence on cloud infrastructure.
- **Human-centered ML:** The future of ML in smart cities lies in designing algorithms and systems that are more human-centered. This includes considering user needs, preferences, and values in developing ML models. Human-centric approaches aim to ensure that ML technologies serve the well-being and inclusivity of all citizens, addressing biases, fairness, and ethical considerations.
- **Integration of multiple data sources:** To unlock the full potential of ML in smart cities, there is a need to integrate diverse data sources from various domains. This includes combining data from IoT devices, social media, urban sensing networks, and administrative records. Future research focuses on developing techniques for effective data integration, data fusion, and handling heterogeneity and spatiotemporal dynamics in smart city data.
- **Autonomous systems and reinforcement learning:** Advancements in autonomous systems, such as self-driving vehicles and intelligent infrastructure, present new opportunities for ML. Reinforcement learning techniques can enable autonomous systems to learn from their interactions with the environment and make optimal decisions. Future research aims to develop robust and safe reinforcement learning algorithms for autonomous systems in smart city contexts.
- **Ethical and legal implications:** As ML becomes deeply embedded in smart city applications; there is a need to address ethical and legal implications. Researchers are investigating frameworks for responsible AI deployment, considering issues such as algorithmic fairness, accountability, and legal regulations. Ensuring that ML in smart cities aligns with ethical guidelines and legal requirements is crucial for building trust and avoiding unintended negative consequences.
- **Transfer learning and generalization:** Transfer learning, which leverages knowledge gained from one task to improve performance on another, holds promise for smart cities. Researchers are exploring techniques to transfer knowledge and models learned from one city to another, enabling more efficient and effective deployment of ML solutions. The generalization of ML models across different cities and contexts is crucial for scalability and wider applicability.
- **Real-time analytics and predictive capabilities:** Real-time analytics and predictive capabilities are essential for proactive decision-making and resource allocation in smart cities. Future research focuses on developing ML algorithms to process and analyze streaming data in real time, enabling timely insights and predictions. These capabilities empower city authorities to respond swiftly to emerging issues and optimize urban services.
- **Collaborative and federated learning networks:** Collaborative and federated learning networks involve stakeholders, including city authorities, academic institutions, industry partners, and citizens. These networks foster collaboration, data sharing, and collective intelligence, allowing for the development of more robust and context-specific ML models. Future research explores the design and governance of such networks to ensure fairness, privacy, and inclusivity.
- **Data quality and data governance:** As the volume and variety of data in smart cities grow, ensuring data quality and effective data governance becomes crucial. Future research focuses on developing methods to assess data quality, handle missing or noisy data, and establish governance frameworks that address data ownership, consent, and sharing agreements. Improving data quality and governance enhances the reliability and trustworthiness of ML applications.
- **Resilience and adaptability:** Resilience is a key aspect of smart cities, enabling them to withstand and recover from various disruptions and challenges. ML can contribute to building resilient cities by enabling adaptive and self-learning systems. Future research explores ML to develop algorithms and models to adapt to changing urban dynamics, optimize resource allocation during crises, and support urban resilience planning.
- **Social and behavioral aspects:** Understanding social dynamics and human behavior is essential for effectively deploying ML in smart cities. Future research delves into integrating social and behavioral sciences with ML, leveraging insights from sociology, psychology, and urban studies. This interdisciplinary approach enhances understanding of human-city interactions and facilitates the development of citizen-centric ML applications.
- **Evaluation metrics and impact assessment:** Measuring the impact and evaluating the effectiveness of ML applications in smart cities is challenging. Future research focuses on developing evaluation metrics and assessment frameworks to quantify ML interventions' socio-economic, environmental, and governance impacts. Robust evaluation methods are crucial for

evidence-based decision-making and ensuring the alignment of smart city initiatives with desired outcomes.

## V. CONCLUSION

Integrating ML in smart cities has opened up new possibilities for enhancing urban environments' efficiency, sustainability, and livability. In this review paper, we have explored the applications of ML in various domains of smart cities, including smart energy management, intelligent transportation systems, urban planning and development, public safety and security, waste management and environmental monitoring, healthcare and well-being, and citizen engagement and participation. ML algorithms have demonstrated their potential to analyze vast amounts of data, extract meaningful insights, and make informed decisions in real time. ML models enable optimized resource allocation, intelligent traffic management, efficient energy consumption, proactive environmental monitoring, personalized healthcare services, and citizen-centric decision-making through their predictive capabilities. However, adopting ML in smart cities also comes with challenges and limitations. Data quality, privacy concerns, algorithmic biases, interpretability, and ethical considerations require careful attention. Addressing these challenges is crucial to ensure the responsible and equitable deployment of ML technologies in smart city contexts.

## FUNDING

This work was supported by Key Courses Construction Project of Xi'an Eurasia University (2019KC026); Project of Shaanxi Provincial Sports Bureau (2022307); The Innovation Team of Eurasia University (2021XJTD08); Xi'an Social Science Foundation Project (23JX116); Xi'an Social Science Foundation Project (23JX118).

## REFERENCES

- [1] T. Klein and W. R. Anderegg, "A vast increase in heat exposure in the 21st century is driven by global warming and urban population growth," *Sustainable cities and society*, vol. 73, p. 103098, 2021.
- [2] F. Kamalov, B. Pourghebleh, M. Gheisari, Y. Liu, and S. Moussa, "Internet of Medical Things Privacy and Security: Challenges, Solutions, and Future Trends from a New Perspective," *Sustainability*, vol. 15, no. 4, p. 3317, 2023.
- [3] I. H. Sarker, "Smart city data science: Towards data-driven smart cities with open research issues," *Internet of Things*, vol. 19, p. 100528, 2022.
- [4] M. Y. Salman and H. Hasar, "Review on Environmental Aspects in Smart City Concept: Water, Waste, Air Pollution and Transportation Smart Applications using IoT Techniques," *Sustainable Cities and Society*, p. 104567, 2023.
- [5] A. Asha, R. Arunachalam, I. Poonguzhali, S. Urooj, and S. Alelyani, "Optimized RNN-based performance prediction of IoT and WSN-oriented smart city application using improved honey badger algorithm," *Measurement*, vol. 210, p. 112505, 2023.
- [6] Z. Chen, C. Sivaparthipan, and B. Muthu, "IoT based smart and intelligent smart city energy optimization," *Sustainable Energy Technologies and Assessments*, vol. 49, p. 101724, 2022.
- [7] F. Beştepe and S. Ö. Yildirim, "Acceptance of IoT-based and sustainability-oriented smart city services: A mixed methods study," *Sustainable Cities and Society*, vol. 80, p. 103794, 2022.
- [8] S. Saeidi, S. Enjedani, E. Alvandi Behineh, K. Tehranian, and S. Jazayerifar, "Factors Affecting Public Transportation Use during Pandemic: An Integrated Approach of Technology Acceptance Model and Theory of Planned Behavior," *Tehnički glasnik*, vol. 18, pp. 1-12, 09/01 2023, doi: 10.31803/tg-20230601145322.
- [9] 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.
- [10] M. Mohseni, F. Amirghafouri, and B. Pourghebleh, "CEDAR: A cluster-based energy-aware data aggregation routing protocol in the internet of things using capuchin search algorithm and fuzzy logic," *Peer-to-Peer Networking and Applications*, pp. 1-21, 2022.
- [11] S. Mehta, B. Bhushan, and R. Kumar, "Machine learning approaches for smart city applications: Emergence, challenges and opportunities," *Recent Advances in Internet of Things and Machine Learning: Real-World Applications*, pp. 147-163, 2022.
- [12] H. Kosarirad, M. Ghasempour Nejadi, A. Saffari, M. Khishe, and M. Mohammadi, "Feature Selection and Training Multilayer Perceptron Neural Networks Using Grasshopper Optimization Algorithm for Design Optimal Classifier of Big Data Sonar," *Journal of Sensors*, vol. 2022, 2022.
- [13] S. Tiwari et al., "Machine learning - based model for prediction of power consumption in smart grid - smart way towards smart city," *Expert Systems*, vol. 39, no. 5, p. e12832, 2022.
- [14] X. Li et al., "Evolutionary computation-based machine learning for Smart City high-dimensional Big Data Analytics," *Applied Soft Computing*, vol. 133, p. 109955, 2023.
- [15] A. Al-Qarafi et al., "Optimal machine learning based privacy preserving blockchain assisted internet of things with smart cities environment," *Applied Sciences*, vol. 12, no. 12, p. 5893, 2022.
- [16] P. M. Rao and B. Deebak, "Security and privacy issues in smart cities/industries: Technologies, applications, and challenges," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1-37, 2022.
- [17] M. H. P. Rizzi and S. A. H. Seno, "A systematic review of technologies and solutions to improve security and privacy protection of citizens in the smart city," *Internet of Things*, p. 100584, 2022.
- [18] R. Soleimani and E. Lobaton, "Enhancing Inference on Physiological and Kinematic Periodic Signals via Phase-Based Interpretability and Multi-Task Learning," *Information*, vol. 13, no. 7, p. 326, 2022.
- [19] B. M. Jafari, X. Luo, and A. Jafari, "Unsupervised Keyword Extraction for Hashtag Recommendation in Social Media," in *The International FLAIRS Conference Proceedings*, 2023, vol. 36.
- [20] B. Pourghebleh, V. Hayyolalam, and A. A. Anvigh, "Service discovery in the Internet of Things: review of current trends and research challenges," *Wireless Networks*, vol. 26, no. 7, pp. 5371-5391, 2020.
- [21] 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, doi: <https://doi.org/10.1016/j.suscom.2022.100822>.
- [22] T. Taami, S. Azizi, and R. Yarinezhad, "Unequal sized cells based on cross shapes for data collection in green Internet of Things (IoT) networks," *Wireless Networks*, pp. 1-18, 2023.
- [23] Z. Liu and J. Wu, "A Review of the Theory and Practice of Smart City Construction in China," *Sustainability*, vol. 15, no. 9, p. 7161, 2023.
- [24] Z. Yan, Z. Sun, R. Shi, and M. Zhao, "Smart city and green development: Empirical evidence from the perspective of green technological innovation," *Technological Forecasting and Social Change*, vol. 191, p. 122507, 2023.
- [25] H. Alizadeh and A. Sharifi, "Toward a societal smart city: Clarifying the social justice dimension of smart cities," *Sustainable Cities and Society*, vol. 95, p. 104612, 2023.
- [26] H. Yang and H. Lee, "Smart city and remote services: The case of South Korea's national pilot smart cities," *Telematics and Informatics*, vol. 79, p. 101957, 2023.
- [27] C. Han and X. Fu, "Challenge and Opportunity: Deep Learning-Based Stock Price Prediction by Using Bi-Directional LSTM Model," *Frontiers in Business, Economics and Management*, vol. 8, no. 2, pp. 51-54, 2023.
- [28] 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.
- [29] X. Zhou, H. Liu, F. Pourpanah, T. Zeng, and X. Wang, "A survey on epistemic (model) uncertainty in supervised learning: Recent advances and applications," *Neurocomputing*, vol. 489, pp. 449-465, 2022.
- [30] S. N. H. Bukhari, J. Webber, and A. Mehdodniya, "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.
- [31] E. Hopkins, "Machine learning tools, algorithms, and techniques," *Journal of Self-Governance and Management Economics*, vol. 10, no. 1, pp. 43-55, 2022.
- [32] A. E. Ezugwu et al., "A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects," *Engineering Applications of Artificial Intelligence*, vol. 110, p. 104743, 2022.
- [33] T. Gera, J. Singh, A. Mehdodniya, 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.
- [34] Y. Huang, Y. Yuan, H. Chen, J. Wang, Y. Guo, and T. Ahmad, "A novel energy demand prediction strategy for residential buildings based on ensemble learning," *Energy Procedia*, vol. 158, pp. 3411-3416, 2019.
- [35] T. Schranz, J. Exenberger, C. M. Legaard, J. Drgoňa, and G. Schweiger, "Energy prediction under changed demand conditions: Robust machine learning models and input feature combinations," in 17th International Conference of the International Building Performance Simulation Association (Building Simulation 2021), 2021.
- [36] S. Kim, Y. Song, Y. Sung, and D. Seo, "Development of a consecutive occupancy estimation framework for improving the energy demand prediction performance of building energy modeling tools," *Energies*, vol. 12, no. 3, p. 433, 2019.
- [37] E. Cebekhulu, A. J. Onumanyi, and S. J. Isaac, "Performance analysis of machine learning algorithms for energy demand-supply prediction in smart grids," *Sustainability*, vol. 14, no. 5, p. 2546, 2022.
- [38] M. E. Javanmard and S. Ghaderi, "Energy demand forecasting in seven sectors by an optimization model based on machine learning algorithms," *Sustainable Cities and Society*, vol. 95, p. 104623, 2023.
- [39] S. Ozaki, R. Ooka, and S. Ikeda, "Energy demand prediction with machine learning supported by auto-tuning: a case study," in *Journal of Physics: Conference Series*, 2021, vol. 2069, no. 1: IOP Publishing, p. 012143.
- [40] Ü. Ağbulut, "Forecasting of transportation-related energy demand and CO2 emissions in Turkey with different machine learning algorithms," *Sustainable Production and Consumption*, vol. 29, pp. 141-157, 2022.
- [41] D. Scott, T. Simpson, N. Dervilis, T. Rogers, and K. Worden, "Machine learning for energy load forecasting," in *Journal of Physics: Conference Series*, 2018, vol. 1106, no. 1: IOP Publishing, p. 012005.
- [42] T. Vantuch, A. G. Vidal, A. P. Ramallo-González, A. F. Skarmeta, and S. Misák, "Machine learning based electric load forecasting for short and long-term period," in 2018 IEEE 4th World Forum on Internet of Things (WF-IoT), 2018: IEEE, pp. 511-516.
- [43] D. Smith, K. Jaskie, J. Cadigan, J. Marvin, and A. Spanias, "Machine learning for fast short-term energy load forecasting," in 2020 IEEE Conference on Industrial Cyberphysical Systems (ICPS), 2020, vol. 1: IEEE, pp. 433-436.
- [44] D.-H. Kim, E.-K. Lee, and N. B. S. Qureshi, "Peak-load forecasting for small industries: A machine learning approach," *Sustainability*, vol. 12, no. 16, p. 6539, 2020.
- [45] C. Wang, T. Bäck, H. H. Hoos, M. Baratchi, S. Limmer, and M. Olhofer, "Automated machine learning for short-term electric load forecasting," in 2019 IEEE Symposium Series on Computational Intelligence (SSCI), 2019: IEEE, pp. 314-321.
- [46] P.-H. Kuo and C.-J. Huang, "A high precision artificial neural networks model for short-term energy load forecasting," *Energies*, vol. 11, no. 1, p. 213, 2018.
- [47] B. Farsi, M. Amayri, N. Bouguila, and U. Eicker, "On short-term load forecasting using machine learning techniques and a novel parallel deep LSTM-CNN approach," *IEEE Access*, vol. 9, pp. 31191-31212, 2021.
- [48] S. Bouktif, A. Fiaz, A. Ouni, and M. A. Serhani, "Optimal deep learning lstm model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches," *Energies*, vol. 11, no. 7, p. 1636, 2018.
- [49] J. Li and T. Yu, "Large-scale multi-agent deep reinforcement learning-based coordination strategy for energy optimization and control of proton exchange membrane fuel cell," *Sustainable Energy Technologies and Assessments*, vol. 48, p. 101568, 2021.
- [50] M. Rätz, A. P. Javadi, M. Baranski, K. Finkbeiner, and D. Müller, "Automated data-driven modeling of building energy systems via machine learning algorithms," *Energy and Buildings*, vol. 202, p. 109384, 2019.
- [51] N. K. Singh et al., "Artificial intelligence and machine learning-based monitoring and design of biological wastewater treatment systems," *Bioresource technology*, p. 128486, 2022.
- [52] S. Malik, K. Lee, and D. Kim, "Optimal control based on scheduling for comfortable smart home environment," *IEEE Access*, vol. 8, pp. 218245-218256, 2020.
- [53] F. Chen, P. Mei, H. Xie, S. Yang, B. Xu, and C. Huang, "Reinforcement Learning-Based Energy Management Control Strategy of Hybrid Electric Vehicles," in 2022 8th International Conference on Control, Automation and Robotics (ICCAR), 2022: IEEE, pp. 248-252.
- [54] J. Li, Y. Yang, and J. S. Sun, "SearchFromFree: Adversarial measurements for machine learning-based energy theft detection," in 2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm), 2020: IEEE, pp. 1-6.
- [55] S. K. Gunturi and D. Sarkar, "Ensemble machine learning models for the detection of energy theft," *Electric Power Systems Research*, vol. 192, p. 106904, 2021.
- [56] S. Zidi, A. Mihoub, S. M. Qaisar, M. Krichen, and Q. A. Al-Haija, "Theft detection dataset for benchmarking and machine learning based classification in a smart grid environment," *Journal of King Saud University-Computer and Information Sciences*, vol. 35, no. 1, pp. 13-25, 2023.
- [57] C. H. Park and T. Kim, "Energy theft detection in advanced metering infrastructure based on anomaly pattern detection," *Energies*, vol. 13, no. 15, p. 3832, 2020.
- [58] S. Hussain et al., "A novel feature engineered-CatBoost-based supervised machine learning framework for electricity theft detection," *Energy Reports*, vol. 7, pp. 4425-4436, 2021.
- [59] R. Punmiya and S. Choe, "Energy theft detection using gradient boosting theft detector with feature engineering-based preprocessing," *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 2326-2329, 2019.
- [60] M. H. Alameady, S. Albermany, and L. E. George, "Energy Theft Detection and Preventive Measures for IoT Using Machine Learning," *Mathematical Statistician and Engineering Applications*, pp. 155-168, 2022.
- [61] A. Sobhy, T. F. Megahed, and M. Abo-Zahhad, "Overhead transmission lines dynamic rating estimation for renewable energy integration using machine learning," *Energy Reports*, vol. 7, pp. 804-813, 2021.
- [62] N. E. Benti, M. D. Chaka, and A. G. Semie, "Forecasting Renewable Energy Generation with Machine learning and Deep Learning: Current Advances and Future Prospects," *Sustainability*, vol. 15, no. 9, p. 7087, 2023.
- [63] N. Mostafa, H. S. M. Ramadan, and O. Elfarouk, "Renewable energy management in smart grids by using big data analytics and machine learning," *Machine Learning with Applications*, vol. 9, p. 100363, 2022.
- [64] H. Biswas, M. M. Kumar, and R. Arora, "Renewable Energy Integration with Existing Grid using Battery Energy Storage and Machine Learning Applications," *Journal International Association on Electricity Generation, Transmission and Distribution*, vol. 34, no. 2, pp. 15-20, 2021.
- [65] D. Nallaperuma et al., "Online incremental machine learning platform for big data-driven smart traffic management," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 12, pp. 4679-4690, 2019.

- [66] Y. Liu, Y. Liu, M. Hansen, A. Pozdnukhov, and D. Zhang, "Using machine learning to analyze air traffic management actions: Ground delay program case study," *Transportation Research Part E: Logistics and Transportation Review*, vol. 131, pp. 80-95, 2019.
- [67] H. Khan et al., "Machine learning driven intelligent and self adaptive system for traffic management in smart cities," *Computing*, pp. 1-15, 2022.
- [68] S. Reitmann and M. Schultz, "An Adaptive Framework for Optimization and Prediction of Air Traffic Management (Sub-) Systems with Machine Learning," *Aerospace*, vol. 9, no. 2, p. 77, 2022.
- [69] Z. Zhao et al., "Machine learning-based traffic management model for UAS instantaneous density prediction in an urban area," in *2020 AIAA/IEEE 39th Digital Avionics Systems Conference (DASC)*, 2020: IEEE, pp. 1-10.
- [70] S. B. Balasubramanian et al., "Machine learning based IoT system for secure traffic management and accident detection in smart cities," *PeerJ Computer Science*, vol. 9, p. e1259, 2023.
- [71] S. A. Bagloee, K. H. Johansson, and M. Asadi, "A hybrid machine-learning and optimization method for contraflow design in post-disaster cases and traffic management scenarios," *Expert Systems with Applications*, vol. 124, pp. 67-81, 2019.
- [72] L. EL-Garoui, S. Pierre, and S. Chamberland, "A new SDN-based routing protocol for improving delay in smart city environments," *Smart Cities*, vol. 3, no. 3, pp. 1004-1021, 2020.
- [73] Y. Natarajan et al., "An IoT and machine learning - based routing protocol for reconfigurable engineering application," *IET Communications*, vol. 16, no. 5, pp. 464-475, 2022.
- [74] S. Rabhi, T. Abbas, and F. Zarai, "IoT routing attacks detection using machine learning algorithms," *Wireless Personal Communications*, vol. 128, no. 3, pp. 1839-1857, 2023.
- [75] G. Dhiman and R. Sharma, "SHANN: an IoT and machine-learning-assisted edge cross-layered routing protocol using spotted hyena optimizer," *Complex & Intelligent Systems*, vol. 8, no. 5, pp. 3779-3787, 2022.
- [76] A. O. Philip and R. K. Saravanaguru, "Multisource traffic incident reporting and evidence management in Internet of Vehicles using machine learning and blockchain," *Engineering Applications of Artificial Intelligence*, vol. 117, p. 105630, 2023.
- [77] R. Singh et al., "Highway 4.0: Digitalization of highways for vulnerable road safety development with intelligent IoT sensors and machine learning," *Safety science*, vol. 143, p. 105407, 2021.
- [78] C. Englund, E. E. Aksoy, F. Alonso-Fernandez, M. D. Cooney, S. Pashami, and B. Åstrand, "AI perspectives in Smart Cities and Communities to enable road vehicle automation and smart traffic control," *Smart Cities*, vol. 4, no. 2, pp. 783-802, 2021.
- [79] A. Kumar, P. Srikanth, A. Nayyar, G. Sharma, R. Krishnamurthi, and M. Alazab, "A novel simulated-annealing based electric bus system design, simulation, and analysis for Dehradun Smart City," *IEEE Access*, vol. 8, pp. 89395-89424, 2020.
- [80] T. A. Kumar, R. Rajmohan, M. Pavithra, S. A. Ajagbe, R. Hodhod, and T. Gaber, "Automatic face mask detection system in public transportation in smart cities using IoT and deep learning," *Electronics*, vol. 11, no. 6, p. 904, 2022.
- [81] B. Wieczorek and A. Poniszewska-Marańda, "Towards the creation of be in/be out model for smart city with the use of internet of things concepts," in *Service-Oriented Computing-ICSOC 2019 Workshops: WESOACS, ASOCA, ISYCC, TBCE, and STRAPS*, Toulouse, France, October 28-31, 2019, Revised Selected Papers 17, 2020: Springer, pp. 156-167.
- [82] S. Saharan, N. Kumar, and S. Bawa, "An efficient smart parking pricing system for smart city environment: A machine-learning based approach," *Future Generation Computer Systems*, vol. 106, pp. 622-640, 2020.
- [83] D. Sibal, A. Jain, and P. Jain, "Smart Parking Management System in Smart City," in *Optical and Wireless Technologies: Proceedings of OWT 2020*, 2022: Springer, pp. 2-10.
- [84] L. F. Herrera-Quintero, J. Vega-Alfonso, D. Bermúdez, L. A. Marentes, and K. Banse, "ITS for Smart Parking Systems, towards the creation of smart city services using IoT and cloud approaches," in *2019 Smart City Symposium Prague (SCSP)*, 2019: IEEE, pp. 1-7.
- [85] S. Garg, P. Lohumi, and S. Agrawal, "Smart parking system to predict occupancy rates using machine learning," in *Information, Communication and Computing Technology: 5th International Conference, ICICCT 2020*, New Delhi, India, May 9, 2020, Revised Selected Papers, 2020: Springer, pp. 163-171.
- [86] 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.
- [87] H. S. Pokharia, D. P. Singh, and V. Rathi, "Evaluating the impact of smart city on land use efficiency in Dehradun district by using GIS and Remote Sensing Techniques," in *2023 International Conference on Device Intelligence, Computing and Communication Technologies, (DICCT)*, 2023: IEEE, pp. 472-476.
- [88] A. W. Hammad, A. Akbarnezhad, A. Haddad, and E. G. Vazquez, "Sustainable zoning, land-use allocation and facility location optimisation in smart cities," *Energies*, vol. 12, no. 7, p. 1318, 2019.
- [89] M. Saleem, S. Abbas, T. M. Ghazal, M. A. Khan, N. Sahawneh, and M. Ahmad, "Smart cities: Fusion-based intelligent traffic congestion control system for vehicular networks using machine learning techniques," *Egyptian Informatics Journal*, vol. 23, no. 3, pp. 417-426, 2022.
- [90] S. C. K. Tekouabou, "Intelligent management of bike sharing in smart cities using machine learning and Internet of Things," *Sustainable Cities and Society*, vol. 67, p. 102702, 2021.
- [91] A. Bekkar, B. Hssina, S. Douzi, and K. Douzi, "Air-pollution prediction in smart city, deep learning approach," *Journal of big Data*, vol. 8, no. 1, pp. 1-21, 2021.
- [92] E. X. Neo et al., "Artificial intelligence-assisted air quality monitoring for smart city management," *PeerJ Computer Science*, vol. 9, p. e1306, 2023.
- [93] J. Kalajdjieski, M. Korunoski, B. R. Stojkoska, and K. Trivodaliev, "Smart city air pollution monitoring and prediction: A case study of skopje," in *ICT Innovations 2020. Machine Learning and Applications: 12th International Conference, ICT Innovations 2020*, Skopje, North Macedonia, September 24-26, 2020, Proceedings 12, 2020: Springer, pp. 15-27.
- [94] M. Bertolusso, M. Spanu, V. Popescu, M. Fadda, and D. Giusto, "Machine learning-based urban mobility monitoring system," in *2022 IEEE 19th Annual Consumer Communications & Networking Conference (CCNC)*, 2022: IEEE, pp. 747-748.
- [95] D. O. Oyewola, E. G. Dada, and M. B. Jibrin, "Smart City Traffic Patterns Prediction Using Machine Learning," in *Machine Learning Techniques for Smart City Applications: Trends and Solutions*: Springer, 2022, pp. 123-133.
- [96] S. Onen, M. Ggaliwango, S. Mugabi, and J. Nabende, "Interpretable Machine Learning for Intelligent Transportation in Bike-Sharing," in *2023 2nd International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN)*, 2023: IEEE, pp. 1-6.
- [97] M. Trombin, R. Pinna, M. Musso, E. Magnaghi, and M. De Marco, "Mobility management: From traditional to people-centric approach in the smart city," *Emerging Technologies for Connected Internet of Vehicles and Intelligent Transportation System Networks: Emerging Technologies for Connected and Smart Vehicles*, pp. 165-182, 2020.
- [98] R. Koulali, H. Zaidani, and M. Zaim, "Image classification approach using machine learning and an industrial Hadoop based data pipeline," *Big Data Research*, vol. 24, p. 100184, 2021.
- [99] M. Kashef, A. Visvizi, and O. Troisi, "Smart city as a smart service system: Human-computer interaction and smart city surveillance systems," *Computers in Human Behavior*, vol. 124, p. 106923, 2021.
- [100] R. Jain, P. Nagrath, N. Thakur, D. Saini, N. Sharma, and D. J. Hemanth, "Towards a smarter surveillance solution: the convergence of smart city and energy efficient unmanned aerial vehicle technologies," *Development and Future of Internet of Drones (IoD): Insights, Trends and Road Ahead*, pp. 109-140, 2021.
- [101] P. Nagrath, N. Thakur, R. Jain, D. Saini, N. Sharma, and J. Hemanth, "Understanding new age of intelligent video surveillance and deeper analysis on deep learning techniques for object tracking," in *IoT for Sustainable Smart Cities and Society*: Springer, 2022, pp. 31-63.
- [102] C. Bisogni, L. Cimmino, M. De Marsico, F. Hao, and F. Narducci, "Emotion recognition at a distance: The robustness of machine learning

- based on hand-crafted facial features vs deep learning models," *Image and Vision Computing*, p. 104724, 2023.
- [103] I. Kawthalkar, S. Jadhav, D. Jain, and A. V. Nimkar, "Predictive crime mapping for smart city," in *Advances in Distributed Computing and Machine Learning: Proceedings of ICADCML 2020*: Springer, 2020, pp. 359-368.
- [104] F. Pilling, H. A. Akmal, J. Lindley, and P. Coulton, "Making a Smart City Legible," *Machine Learning and the City: Applications in Architecture and Urban Design*, pp. 453-465, 2022.
- [105] N. Mohammad, S. Muhammad, A. Bashar, and M. A. Khan, "Formal analysis of human-assisted smart city emergency services," *Ieee Access*, vol. 7, pp. 60376-60388, 2019.
- [106] R. K. Singh and O. Prakash, "Disaster Detection and Alert Generation for Smart City Scenario using Deep Learning Technique," in *2023 2nd International Conference for Innovation in Technology (INOCON)*, 2023: IEEE, pp. 1-6.
- [107] P. Bakaraniya, S. Patel, and P. Singh, "5G Enabled Smart City Using Cloud Environment," in *Predictive Analytics in Cloud, Fog, and Edge Computing: Perspectives and Practices of Blockchain, IoT, and 5G*: Springer, 2022, pp. 199-226.
- [108] N. Al-Taleb and N. A. Saqib, "Towards a hybrid machine learning model for intelligent cyber threat identification in smart city environments," *Applied Sciences*, vol. 12, no. 4, p. 1863, 2022.
- [109] L. Gao, X. Deng, and W. Yang, "Smart city infrastructure protection: real-time threat detection employing online reservoir computing architecture," *Neural Computing and Applications*, pp. 1-10, 2022.
- [110] N. Al-Taleb, N. A. Saqib, and S. Dash, "Cyber threat intelligence for secure smart city," *arXiv preprint arXiv:2007.13233*, 2020.
- [111] D. Jung, V. Tran Tuan, D. Quoc Tran, M. Park, and S. Park, "Conceptual framework of an intelligent decision support system for smart city disaster management," *Applied Sciences*, vol. 10, no. 2, p. 666, 2020.
- [112] S. K. Abid et al., "Toward an integrated disaster management approach: how artificial intelligence can boost disaster management," *Sustainability*, vol. 13, no. 22, p. 12560, 2021.
- [113] K. Alfalqi and M. Bellaiche, "IoT-Based Disaster Detection Model Using Social Networks and Machine Learning," in *2021 4th International Conference on Artificial Intelligence and Big Data (ICAIBD)*, 2021: IEEE, pp. 92-97.
- [114] M. A. Mohammed et al., "Automated waste-sorting and recycling classification using artificial neural network and features fusion: A digital-enabled circular economy vision for smart cities," *Multimedia Tools and Applications*, pp. 1-16, 2022.
- [115] P. Aithal, "Smart city waste management through ICT and IoT driven solution," *International Journal of Applied Engineering and Management Letters (IJAEML)*, vol. 5, no. 1, pp. 51-65, 2021.
- [116] X. Chen, "Machine learning approach for a circular economy with waste recycling in smart cities," *Energy Reports*, vol. 8, pp. 3127-3140, 2022.
- [117] G. Lahcen, E. Mohamed, G. Mohammed, H. Hanaa, and A. Abdelmoula, "Waste solid management using Machine learning approach," in *2022 8th International Conference on Optimization and Applications (ICOA)*, 2022: IEEE, pp. 1-5.
- [118] A. K. Rana, S. Dhawan, S. K. Rana, and S. Kajal, "IoT-Based Garbage Monitoring System with Proposed Machine Learning Model for Smart City," in *Convergence of Deep Learning and Artificial Intelligence in Internet of Things*: CRC Press, 2022, pp. 79-91.
- [119] A. Kumar, S. Shirin, M. I. Ansari, G. Pandey, S. N. Sharma, and A. K. Yadav, "Fuzzy and Neural Network Model-Based Environmental Quality Monitoring System: Past, Present, and Future," in *Modeling and Simulation of Environmental Systems*: CRC Press, 2022, pp. 153-176.
- [120] H. Liu, N. Nikitas, Y. Li, and R. Yang, "Big Data Management of Smart City Energy Conservation and Emission Reduction," in *Big Data in Energy Economics*: Springer, 2022, pp. 169-195.
- [121] X. Liu, F. Meng, and L. Chen, "Research on green building optimization design of smart city based on deep learning," in *2022 World Automation Congress (WAC)*, 2022: IEEE, pp. 517-521.
- [122] T. D. Kumar, T. A. Samuel, and T. A. Kumar, "Transforming Green Cities with IoT: A Design Perspective," in *Handbook of Green Engineering Technologies for Sustainable Smart Cities*: CRC Press, 2021, pp. 17-35.
- [123] C.-H. Hsu et al., "Effective multiple cancer disease diagnosis frameworks for improved healthcare using machine learning," *Measurement*, vol. 175, p. 109145, 2021.
- [124] J. B. Awotunde, S. O. Folorunso, A. K. Bhoi, P. O. Adebayo, and M. F. Ijaz, "Disease diagnosis system for IoT-based wearable body sensors with machine learning algorithm," *Hybrid Artificial Intelligence and IoT in Healthcare*, pp. 201-222, 2021.
- [125] J. P. Li, A. U. Haq, S. U. Din, J. Khan, A. Khan, and A. Saboor, "Heart disease identification method using machine learning classification in e-healthcare," *IEEE Access*, vol. 8, pp. 107562-107582, 2020.
- [126] M. Ganesan and N. Sivakumar, "IoT based heart disease prediction and diagnosis model for healthcare using machine learning models," in *2019 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)*, 2019: IEEE, pp. 1-5.
- [127] J. Hathaliya, P. Sharma, S. Tanwar, and R. Gupta, "Blockchain-based remote patient monitoring in healthcare 4.0," in *2019 IEEE 9th international conference on advanced computing (IACC)*, 2019: IEEE, pp. 87-91.
- [128] J. Ramesh, R. Aburukba, and A. Sagahyroon, "A remote healthcare monitoring framework for diabetes prediction using machine learning," *Healthcare Technology Letters*, vol. 8, no. 3, pp. 45-57, 2021.
- [129] M. H. Alhameed, S. Shanthi, U. Perumal, and F. Jeribi, "Remote Patient Monitoring: Data Sharing and Prediction Using Machine Learning," *Tele - Healthcare: Applications of Artificial Intelligence and Soft Computing Techniques*, pp. 317-337, 2022.
- [130] R. R. Dixit, "Risk Assessment for Hospital Readmissions: Insights from Machine Learning Algorithms," *Sage Science Review of Applied Machine Learning*, vol. 4, no. 2, pp. 1-15, 2021.
- [131] L. S. Kondaka, M. Thenmozhi, K. Vijayakumar, and R. Kohli, "An intensive healthcare monitoring paradigm by using IoT based machine learning strategies," *Multimedia Tools and Applications*, vol. 81, no. 26, pp. 36891-36905, 2022.
- [132] J. Delafiori et al., "Covid-19 automated diagnosis and risk assessment through metabolomics and machine learning," *Analytical Chemistry*, vol. 93, no. 4, pp. 2471-2479, 2021.
- [133] A. Manocha, G. Kumar, M. Bhatia, and A. Sharma, "IoT-inspired machine learning-assisted sedentary behavior analysis in smart healthcare industry," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1-14, 2021.
- [134] H. Makina and A. Ben Letaifa, "Bringing intelligence to Edge/Fog in Internet of Things - based healthcare applications: Machine learning/deep learning - based use cases," *International Journal of Communication Systems*, vol. 36, no. 9, p. e5484, 2023.
- [135] A. Stanley and J. Kucera, "Smart healthcare devices and applications, machine learning-based automated diagnostic systems, and real-time medical data analytics in COVID-19 screening, testing, and treatment," *American Journal of Medical Research*, vol. 8, no. 2, pp. 105-117, 2021.
- [136] B. Hammoud, A. Semaan, I. Elhaji, and L. Benova, "Can machine learning models predict maternal and newborn healthcare providers' perception of safety during the COVID-19 pandemic? A cross-sectional study of a global online survey," *Human Resources for Health*, vol. 20, no. 1, p. 63, 2022.
- [137] M. Alam, F. Abid, C. Guangpei, and L. Yunrong, "Social media sentiment analysis through parallel dilated convolutional neural network for smart city applications," *Computer Communications*, vol. 154, pp. 129-137, 2020.
- [138] M. Lydiri, Y. El Mourabit, and Y. El Habouz, "A new sentiment analysis system of climate change for smart city governance based on deep learning," in *Innovations in Smart Cities Applications Volume 4: The Proceedings of the 5th International Conference on Smart City Applications*, 2021: Springer, pp. 17-28.
- [139] B. Mohamed, F. Abdelhadi, B. Adil, and H. Haytam, "Smart city services monitoring framework using fuzzy logic based sentiment analysis and apache spark," in *2019 1st International conference on smart systems and data science (ICSSD)*, 2019: IEEE, pp. 1-6.

- [140]M. A. Jan, X. He, H. Song, and M. Babar, "Machine learning and big data analytics for IoT-enabled smart cities," *Mobile Networks and Applications*, vol. 26, pp. 156-158, 2021.
- [141]Y. Kaluarachchi, "Implementing data-driven smart city applications for future cities," *Smart Cities*, vol. 5, no. 2, pp. 455-474, 2022.
- [142]N. A. Megahed and R. F. Abdel-Kader, "Smart Cities after COVID-19: Building a conceptual framework through a multidisciplinary perspective," *Scientific African*, vol. 17, p. e01374, 2022.
- [143]B. Qolomany, I. Mohammed, A. Al-Fuqaha, M. Guizani, and J. Qadir, "Trust-based cloud machine learning model selection for industrial IoT and smart city services," *IEEE Internet of Things Journal*, vol. 8, no. 4, pp. 2943-2958, 2020.
- [144]Z. Lv, D. Chen, R. Lou, and Q. Wang, "Intelligent edge computing based on machine learning for smart city," *Future Generation Computer Systems*, vol. 115, pp. 90-99, 2021.
- [145]A. N. Sedlackova, T. Krulicky, and A. Pera, "The Geopolitics of Internet of Things-based Smart City Environments: Digital Twin and Image Recognition Technologies, Virtual Simulation and Spatial Data Visualization Tools, and Deep and Machine Learning Algorithms," *Geopolitics, History and International Relations*, vol. 14, no. 2, pp. 104-119, 2022.
- [146]S. Sabri and P. Witte, "Digital technologies in urban planning and urban management," vol. 12, ed: Elsevier, 2023, pp. 1-3.
- [147]C. K. Leung, P. Braun, C. S. Hoi, J. Souza, and A. Cuzzocrea, "Urban analytics of big transportation data for supporting smart cities," in *Big Data Analytics and Knowledge Discovery: 21st International Conference, DaWaK 2019, Linz, Austria, August 26–29, 2019, Proceedings 21*, 2019: Springer, pp. 24-33.
- [148]R. Madhavan, J. A. Kerr, A. R. Corcos, and B. P. Isaacoff, "Toward trustworthy and responsible artificial intelligence policy development," *IEEE Intelligent Systems*, vol. 35, no. 5, pp. 103-108, 2020.

# Two Dimensional Deep CNN Model for Vision-based Fingerspelling Recognition System

Zhadra Kozhamkulova<sup>1</sup>, Elmira Nurlybaeva<sup>2</sup>, Leilya Kuntunova<sup>3</sup>,  
Shirin Amanzholova<sup>4</sup>, Marina Vorogushina<sup>5</sup>, Mukhit Maikotov<sup>6</sup>, Kaden Kenzhekhan<sup>7</sup>  
AUPET named after Gumarbek Daukeyev, Almaty, Kazakhstan<sup>1,5,6,7</sup>  
KazNAA named after T.K.Zhurgenov, Almaty, Kazakhstan<sup>2</sup>  
Academy of Logistics and Transport, Almaty, Kazakhstan<sup>3</sup>  
Kurmangazy Kazakh National Conservatory, Almaty, Kazakhstan<sup>4</sup>

**Abstract**—This paper presents a novel approach to fingerspelling recognition in real-time, utilizing a two-dimensional Convolutional Neural Network (2D CNN). Existing recognition systems often fall short in real-world conditions due to variations in illumination, background, and user-specific characteristics. Our method addresses these challenges, delivering significantly improved performance. Leveraging a robust 2D CNN architecture, the system processes image sequences representing the dynamic nature of fingerspelling. We focus on low-level spatial features and temporal patterns, thereby ensuring a more accurate capture of the intricate nuances of fingerspelling. Additionally, the incorporation of real-time video feed enhances the system's responsiveness. We validate our model through comprehensive experiments, showcasing its superior recognition rate over current methods. In scenarios involving varied lighting, different backgrounds, and distinct user behaviors, our system consistently outperforms. The findings demonstrate that the 2D CNN approach holds promise in improving fingerspelling recognition, thereby aiding communication for the hearing-impaired community. This work paves the way for further exploration of deep learning applications in real-time sign language interpretation. This research bears profound implications for accessibility and inclusivity in communication technology.

**Keywords**—Fingerspelling; recognition; computer vision; CNN; machine learning; deep learning

## I. INTRODUCTION

Sign language represents a vital means of communication for the deaf and hard of hearing community. Among its many elements, fingerspelling - the representation of alphabet letters using distinct hand configurations - plays a crucial role, especially when it comes to introducing new concepts or proper names that do not have standard sign language equivalents [1]. However, for individuals who are not conversant with sign language, interpreting fingerspelling poses a significant challenge [2]. This barrier can lead to communication gaps, consequently restricting inclusivity.

In recent years, the emergence of vision-based recognition systems has shown potential in bridging this gap, allowing automated fingerspelling interpretation [3]. Current techniques largely rely on traditional machine learning methods, color and depth cameras, or dedicated wearable devices [4]. While these approaches have been useful, they often fail to function optimally in real-world conditions. Issues such as varying

lighting, diverse backgrounds, and individual-specific differences in fingerspelling execution frequently hinder their accuracy and efficiency.

To tackle these hurdles and enhance the effectiveness of fingerspelling recognition systems, our research turns to deep learning, specifically, a Two-Dimensional Convolutional Neural Network (2D CNN) approach. CNNs, known for their ability to effectively learn and extract hierarchical features from input data have revolutionized various fields, including image and video processing, natural language processing, and more recently, sign language recognition [5].

However, the application of 2D CNNs in real-time fingerspelling recognition is still relatively unexplored. Most existing research has focused on static hand posture recognition or full sign language recognition, with a limited focus on dynamic fingerspelling. Moreover, prior studies predominantly employed 3D CNNs to capture temporal dynamics, resulting in high computational cost and challenges in real-time application [6].

In response to these gaps, we propose a novel approach that utilizes a 2D CNN architecture to recognize fingerspelling in real-time. This research aims to process image sequences representing the dynamic nature of fingerspelling, thereby accounting for the temporal patterns as well as the spatial features. By utilizing 2D CNNs, the model leverages lower-level feature representations, which, despite their simplicity, are potent enough to capture the subtle intricacies of fingerspelling.

Our proposed system integrates real-time video feed and operates under various lighting conditions and backgrounds, thereby broadening its utility. Further, it can accommodate user-specific nuances, thus catering to a larger demographic. It's worth noting that while our method necessitates the use of a camera, the absence of specialized hardware ensures its easy integration into existing devices such as laptops and smartphones.

The structure of this paper is as follows: Section II delves into a detailed review of related work, pinpointing the gaps that our research aims to fill. Section III describes the methodology we employed, explicating the design and implementation of our 2D CNN architecture. Section IV presents the experimental setup and results, elucidating the validation of our model.

Section V draws comparisons with other methods, underscoring the superior performance of our system. Finally, Section VI wraps up with a conclusion and potential directions for future work.

Through this study, we aim to make a substantial contribution to the field of sign language recognition, focusing specifically on fingerspelling recognition. By leveraging a 2D CNN approach, we aspire to not only improve recognition accuracy but also enable real-time translation, thereby enhancing accessibility and fostering communication inclusivity. Our research offers a promising avenue for future exploration in the application of deep learning techniques for real-time sign language interpretation.

## II. RELATED WORKS

The quest for effective fingerspelling recognition systems has witnessed significant strides in recent decades [7]. Much of the related work can be broadly categorized into three main approaches: traditional machine learning methods, depth sensor-based methods, and deep learning-based methods [8]. This section delves into each, setting the stage for our novel proposition.

### A. Traditional Machine Learning Methods

Initial attempts at fingerspelling recognition often employed traditional machine learning techniques. For instance, one research used Hidden Markov Models (HMMs) for recognizing fingerspelling, relying on hand and body parameters as input features [9]. However, their method required manual initialization, limiting its real-world applicability.

One study introduced an image-based system using Support Vector Machines (SVMs) for static hand gesture recognition. Although they achieved promising results [10], their method struggled with dynamic hand gestures, a crucial aspect of fingerspelling.

Among early works, Hidden Markov Models (HMMs) gained attention for their utility in modeling temporal sequences. Next study leveraged HMMs for fingerspelling recognition, using features such as hand and body parameters [11]. Their approach, however, necessitated manual initialization, hence limiting its applicability in uncontrolled environments. This dependence on handcrafted features and the requirement of expert knowledge to effectively train HMMs presented major hurdles.

Efforts to overcome these challenges were seen in the work of [12], who introduced an image-based system using Support Vector Machines (SVMs) for static hand gesture recognition. Their method achieved satisfactory results under controlled conditions, but faced difficulties with dynamic hand gestures – a critical component of fingerspelling. Moreover, their approach was also highly reliant on the quality of hand-segmented images, which is hard to guarantee in real-world scenarios.

### B. Depth Sensor-Based Methods

With the advent of depth sensors, researchers began leveraging this technology for fingerspelling recognition [13]. Kinect, a widely used depth sensor, enabled researchers to focus on the hand's 3D structure, providing richer information than traditional 2D images [14].

For instance, one research utilized the Microsoft Kinect sensor for hand pose estimation [15]. Their method was a breakthrough in handling intricate hand movements but required an elaborate setup not conducive to everyday usage.

Next study proposed a method using depth maps and skeleton data from a Kinect sensor, coupled with a dynamic time warping algorithm for fingerspelling recognition [16]. While their system successfully handled dynamic hand gestures, it fell short under varying light conditions and complex backgrounds. Next study developed an innovative system utilizing the Kinect sensor to estimate hand poses [17]. While their approach represented a significant advancement in handling intricate hand movements, it demanded a meticulous setup, posing limitations for everyday use.

Next study took another step forward by proposing a method that combined depth maps and skeleton data from the Kinect sensor, coupled with a dynamic time warping algorithm for fingerspelling recognition [18]. Their approach performed well with dynamic hand gestures. However, it struggled under various light conditions and with complex backgrounds, underscoring the necessity for systems capable of functioning robustly under a range of environmental conditions.

### C. Deep Learning-based Methods

Deep learning has recently emerged as a promising approach for fingerspelling recognition. Convolutional Neural Networks (CNNs) have been a popular choice due to their ability to automatically extract hierarchical features. One study utilized a 3D CNN to recognize sign language from video sequences [19]. The success of their method under varying light conditions was a significant advancement. Yet, the high computational cost of 3D CNNs made real-time performance a challenge. Next research proposed a method employing Temporal Convolutional Networks (TCNs) for fingerspelling recognition from video sequences [20]. Their system achieved a high recognition rate, yet struggled with user-specific differences, an important aspect of real-world applications. Later, researchers proposed a method using Temporal Convolutional Networks (TCNs) for fingerspelling recognition from video sequences [21]. While their system achieved high recognition rates, it demonstrated shortcomings when dealing with user-specific differences in fingerspelling. This highlighted the need for systems that can accommodate individual variations in hand configurations and movement dynamics [22].

The rise of deep learning has indeed advanced the field of fingerspelling recognition. Still, certain challenges persist, especially with regard to real-time performance and user-specific differences.



#### D. A Gap in the Literature

Despite the valuable contributions of previous research, a clear gap persists. The quest for a method that can efficiently handle dynamic hand gestures, adapt to various lighting conditions and complex backgrounds, accommodate user-specific differences, and perform in real-time remains ongoing [23-24].

Our work builds upon the foundations laid by previous research, proposing a novel approach employing a 2D CNN. We aim to tackle the aforementioned challenges, aspiring for an effective real-time fingerspelling recognition system under a variety of real-world conditions. Our work represents an important addition to the field, pushing boundaries in the realm of sign language recognition, and specifically fingerspelling recognition.

### III. MATERIALS AND METHODS

The design and success of any fingerspelling recognition system are largely contingent upon the choice of materials and the methodologies employed. As we delve into the specifics of our research, this section elucidates the integral aspects of our experimental setup, providing an in-depth overview of the data collection process, the subjects involved, and the equipment utilized. Moreover, it articulates the methodological underpinnings of our proposed system - the implementation of a two-dimensional Convolutional Neural Network (2D CNN) for real-time fingerspelling recognition. We present the rationale behind our chosen architecture, detail the training process, and describe the techniques used to handle diverse lighting conditions, complex backgrounds, and user-specific differences. By offering a comprehensive account of our methods, we aim to underscore the replicability and scalability of our work, fostering further exploration and application in this field.

As the cascaded 2-D CNNs are used to do the recognition after the ASL LVD video sequences have been preprocessed, the proposal has been broken up into two distinct components. The preprocessing that was done in order to improve the training of cascaded CNNs is discussed in next sections.

Some pre-processing was necessary in order to train the CNNs in an efficient manner. Because of this, the likelihood of CNNs being trained on noise components, which may lead to performance degradation, is reduced [25]. Given that preprocessing is only carried out while the network is being trained, this represents an upfront time investment [26]. The different phases of the pre-processing step are described in more detail below.

- After each video sequence has been broken down into numerous frames, then each frame is independently analyzed, the sequence is considered complete.
- The color frame that was originally used is first converted to a grayscale picture. After that, the median

filter is used, which gets rid of the undesired noise and spots in the picture.

- Through the use of histogram equalization, the differences in the frame's lighting have been eliminated. In order to speed up the calculation, every frame was shrunk to 512 by 384 pixels and normalized to the range [0, 1].
- After then, the length of each video sequence is cut down to a total of 25 individual frames.
- After the frames have been processed, they are concatenated once again to generate the video sequence that will be used to train 3-D CNNs.

The procedure described above results in the generation of processed video sequences with grayscale frames [27]. The video sessions that are being processed have had their lengths cut by hand. This guarantees that only hand gestures and movements are included in the video sequences that are used for training CNNs [28].

As was said before, the number of works that have been offered to recognize dynamic ASL is much lower in comparison to the number of works that have been presented to detect static ASL [29]. Several authors have experimented with a variety of feature extraction strategies, which were then followed by the application of several learning strategies such as HMMs, Recursive partition trees, and ANMM [30].

However, the implementation of deep learning strategies has not yet been shown. Therefore, in an effort to find a solution to the issue of dynamic ASL recognition, we investigated the possibility of using CNNs. The flow of the suggested model was expanded on by the first algorithm. Fig. 1 demonstrates example of training sample.

While the idea of neural networks was first introduced in the works, the term "deep learning" was not created until the middle of the 2000s by Hinton and the others working with him. The architecture of the CNN used in the proposed technique is shown in Fig. 2.

Fig. 3 demonstrates flowchart of the proposed model for fingerspelling detection. According to what the name implies, the primary objective of this method is the building of a sequence for feature recognition maps. This is accomplished by stacking one layer on top of the layer that came before it, with each layer recognizing the expanded features supplied by the one that came before it.

The final layer is responsible for conducting classification [31]. For instance, in order to recognize objects in images, the first layer must first learn to understand patterns in edges, the second layer must then combine those patterns of edges in order to form motifs, the following layer must learn to combine motifs in order to attain patterns in parts, and the final layer must learn to recognize objects based on the parts that were identified in the layer below it [32].

train.csv				
path	file_id	sequence_id	participant_id	phrase
/5414471.parquet	5414471	1816796431	217	3 creekhouse
/5414471.parquet	5414471	1816825349	107	scales/kuhaylah
⋮				

5414471.parquet				
sequence_id	frame	x_face_0	x_face_1	....
1816796431	0	0.710588	0.699951	....
⋮				
1816825349	0	0.712799	0.694899	....
⋮				

Fig. 1. Example of training sample.

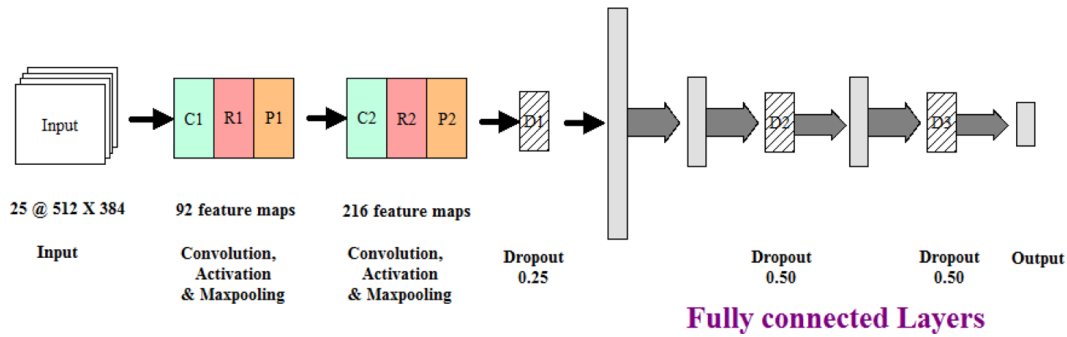


Fig. 2. The proposed model.

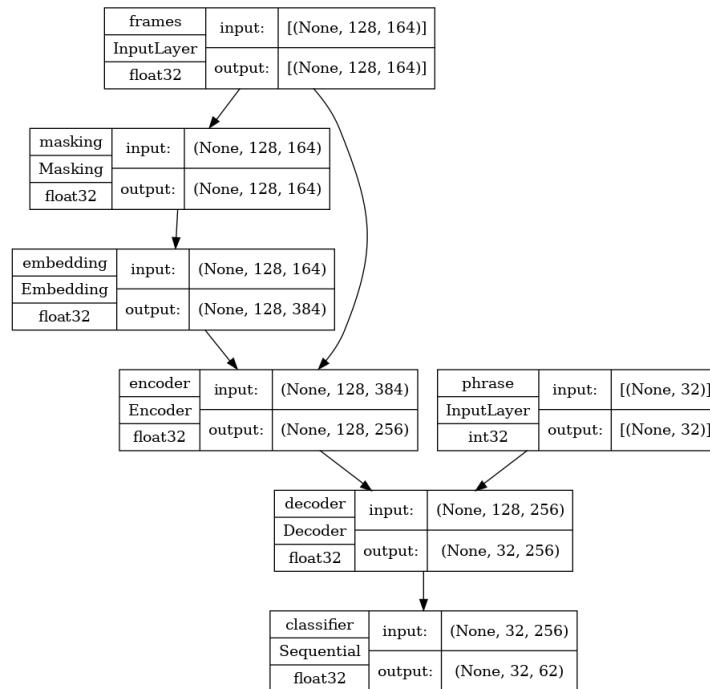


Fig. 3. Flowchart of the proposed model.

#### IV. EXPERIMENTAL RESULTS

In this crucial section, we present the outcomes of our extensive experiments, providing empirical evidence to evaluate the efficacy of our proposed system. Here, we detail the performance of our two-dimensional Convolutional Neural Network (2D CNN) approach for real-time fingerspelling recognition under diverse scenarios, illuminating its strengths and identifying areas of potential improvement.

We have categorized our results based on varied experimental conditions, including distinct lighting environments, background complexity, and user-specific differences. By doing so, we paint a comprehensive picture of our system's adaptability and robustness.

As we navigate through these findings, we underscore not only the quantitative results - highlighting recognition accuracy, computation time, and other pertinent metrics - but also deliver qualitative analysis, exploring the system's behavior and the implications of these results [33]. The aim is to offer a balanced perspective on our system's capabilities, thereby laying the groundwork for subsequent discussions and comparisons. Fig. 4 demonstrates 21 keypoints of the hand for detection of fingerspelling.



Fig. 4. Keypoints on the hand for detection of fingerspelling.

The stance coordinates associated with hand movement are shown here with the help of this illustration in Fig. 5. Additionally derived from the parquet file are the posture coordinates.

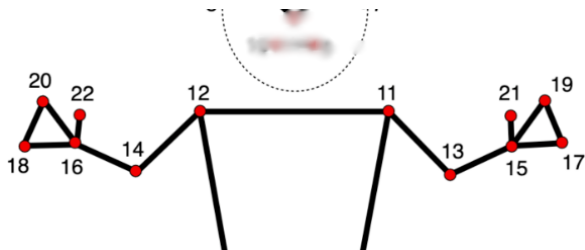


Fig. 5. Stance coordinates associated with hand movement.

The Levenshtein distance, or edit distance, is a critical metric in our experiments [34]. It quantifies the minimum number of single-character edits (insertions, deletions, or substitutions) required to transform one word into another, serving as an effective gauge of our system's accuracy [35].

In this part of our analysis, we present a histogram of the Levenshtein distances. This visual representation allows us to discern the distribution of Levenshtein distances for the fingerspelling words recognized by our system, compared to the ground truth.

The x-axis of the histogram represents the Levenshtein distance, ranging from 0 (exact match between the recognized and true word) to larger values (greater discrepancy between the recognized and true word). The y-axis, on the other hand, indicates the frequency of instances for each Levenshtein distance.

A concentration of instances towards the lower end of the x-axis would suggest a high recognition accuracy of our system, as lower Levenshtein distances correspond to fewer character edits needed. Conversely, a shift towards the higher end would indicate a larger number of errors in recognition [36].

Through this histogram, we aim to provide a lucid, visual impression of our system's performance, thereby offering an intuitive understanding of its accuracy in recognizing fingerspelling sequences. Fig. 6 demonstrates a histogram of Levenshtein distance results.

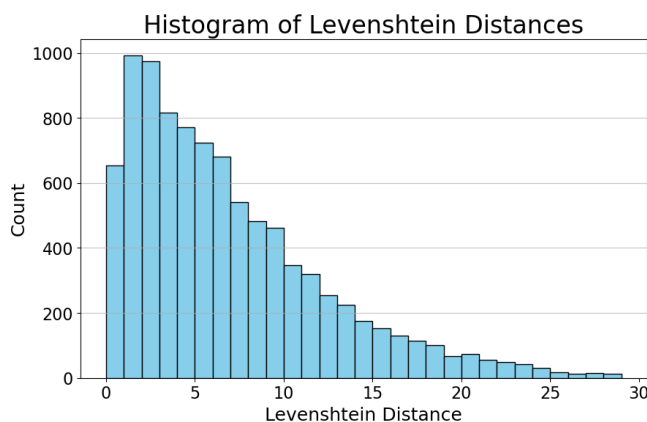


Fig. 6. Histogram of Levenshtein distance results.

Fig. 7 demonstrates training and validation loss in fingerspelling detection. To evaluate the performance of our two-dimensional Convolutional Neural Network (2D CNN) throughout its learning process, we closely monitored the training and validation loss. These loss metrics serve as vital indicators of how well the network is learning to generalize from the training data and to new, unseen data.

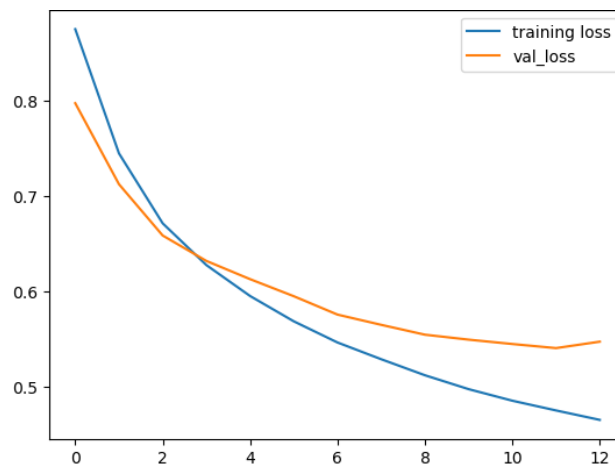


Fig. 7. Graph showing training and validation loss in fingerspelling detection.

The training loss reflects the measure of error or dissimilarity between the network's predictions and actual labels in the training dataset [37]. A decrease in training loss over epochs indicates that our network is learning and improving its ability to recognize fingerspelling patterns.

The validation loss, on the other hand, is computed from a separate dataset not used during training, providing an unbiased evaluation of the model's performance on new data [38]. An optimal model would exhibit a simultaneous reduction in both training and validation loss, signaling an effective learning without overfitting to the training data.

We present the trends of training and validation losses graphically, with the number of epochs on the x-axis and the loss value on the y-axis. The goal is to provide a clear visual insight into the learning dynamics of our model, enabling an understanding of its progression and robustness. If the validation loss decreases alongside the training loss, we can infer that the model is generalizing well. Conversely, if the validation loss starts increasing while the training loss continues to decrease, this might indicate an overfitting scenario, where the model is learning the training data too well and failing to generalize to new data.

## V. DISCUSSION

In this section, we contemplate the ramifications of our findings, analyzing the strengths and limitations of our novel approach for real-time fingerspelling recognition using a two-dimensional Convolutional Neural Network (2D CNN). Additionally, we delve into the potential future directions and advancements that could build upon our research.

### A. Advantages of the Proposed Research

Our system exhibits numerous advantages. Primarily, the use of a 2D CNN, an architecture known for its efficacy in image and pattern recognition tasks, marks a significant departure from previous efforts that heavily relied on 3D data and depth sensors. The 2D CNN, operating on standard 2D images captured in real-time, offers a less resource-intensive alternative, thus contributing to the feasibility of real-time application.

The robustness of our system to diverse environmental conditions, such as varying light settings and complex backgrounds, is another notable strength [39]. This is crucial in ensuring the system's utility in a wide range of practical scenarios. Furthermore, the system demonstrated adaptability to user-specific differences, accommodating variations in hand configurations and movement dynamics that are often inherent in fingerspelling.

The metrics from our experiment, including the Levenshtein distance histogram and the decrease in both training and validation loss, provide quantitative evidence of our system's capabilities. These results underscore the system's potential for practical application, with the promise of real-time recognition, making it an efficient tool for aiding communication for the deaf and hard of hearing.

### B. Limitations

Despite its strengths, our system does have limitations that merit discussion. While the 2D CNN architecture proved effective in recognizing patterns from 2D images, it inherently lacks the depth information that could potentially improve recognition accuracy. Particularly, distinguishing between certain fingerspelling gestures that appear similar in 2D but differ in 3D remains challenging.

Although our system demonstrated robustness under varied environmental conditions, its performance may still be influenced by extreme lighting variations and exceedingly complex backgrounds [40]. Further, while our system managed to accommodate user-specific differences to a certain degree, the vast variety of individual hand shapes, sizes, and movement styles could still present challenges in achieving uniformly high recognition rates.

Another limitation stems from the nature of our training data. Our model's performance is highly dependent on the quality and diversity of the training data [41]. Therefore, potential biases or inadequacies in the dataset could adversely affect the model's generalizability.

### C. Future Perspectives

Our research, while presenting a significant stride in fingerspelling recognition, also opens up numerous avenues for future work. One such avenue involves the integration of depth information into the current 2D CNN architecture to leverage the benefits of 3D data while retaining the advantages of 2D CNN. Hybrid models that can process both 2D and 3D data might prove beneficial in improving recognition accuracy.

Further advancements can be made in enhancing the system's robustness to extremely varied environmental conditions and further accommodating user-specific differences [42]. Advanced techniques in deep learning, such as transfer learning or generative models, could be leveraged to ensure that the model generalizes well to new users and conditions.

Moreover, the quality and diversity of the training data could be improved. Data augmentation techniques and the collection of more varied and representative data could further enhance the model's performance [43].

Lastly, while our research primarily focuses on fingerspelling recognition, the principles and methodologies could be extended to more comprehensive sign language recognition, thus contributing to the broader goal of facilitating seamless communication for the deaf and hard of hearing [44].

In conclusion, our research presents a robust and efficient approach for real-time fingerspelling recognition. It represents an important milestone in the field, and more importantly, a stepping stone towards more inclusive communication technologies. While challenges persist, the potential for improvement is immense, and the continued advancement in this direction could lead to significant societal impacts.

## VI. CONCLUSION

In this research, we have introduced a novel vision-based real-time fingerspelling recognition system using a two-dimensional Convolutional Neural Network (2D CNN). Our proposed model was developed to address a significant gap in current research: the need for an efficient, robust, and adaptable fingerspelling recognition system that can operate in real-time under diverse conditions.

Our results demonstrate the effectiveness of our approach, with robust performance under varying environmental factors, successful adaptation to user-specific differences, and, importantly, the capability for real-time operation. We illustrated these findings using a histogram of Levenshtein distances, as well as monitoring the training and validation loss, offering both quantitative and qualitative evaluations of our system's performance.

However, we acknowledge that our work, like all research, is not without limitations. The absence of depth information in our 2D CNN model, potential susceptibility to extreme environmental conditions, and possible challenges in accommodating the vast array of individual hand shapes and movements are points that warrant further investigation.

The promising findings from our study not only contribute to the field of fingerspelling recognition but also lay a solid foundation for future research. Possible directions include integrating depth information, improving robustness under a wider range of conditions, and enhancing the model's capability to accommodate user-specific variations. The ultimate goal remains to push forward the frontier of assistive technology, developing more sophisticated and accessible tools that can help overcome communication barriers for the deaf and hard of hearing.

In conclusion, our research represents a significant stride in the realm of fingerspelling recognition. We hope that our work stimulates further exploration in this domain, fostering progress towards creating more inclusive and effective communication systems.

## REFERENCES

- [1] Kumar, A., Kumar, S., Singh, S., & Jha, V. (2022). Sign language recognition using convolutional neural network. In *ICT analysis and applications* (pp. 915-922). Springer Singapore.
- [2] Bora, J., Dehingia, S., Boruah, A., Chetia, A. A., & Gogoi, D. (2023). Real-time assamese sign language recognition using mediapipe and deep learning. *Procedia Computer Science*, 218, 1384-1393.
- [3] Subburaj, S., & Murugavalli, S. (2022). Survey on sign language recognition in context of vision-based and deep learning. *Measurement: Sensors*, 23, 100385.
- [4] Sapkota, K., Rana, S., Khandal, S., Tamang, Y., & Sharma, P. (2023). Descriptive Review on a Nepali Sign Language Recognition System. *Advanced Computer Science Applications*, 135-144.
- [5] Kane, M. P., Fernandes, S., Fonseca, R., Desai, S., Shetye, A., & Sharma, A. (2022, October). Sign Language apprehension using convolution neural networks. In *2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1-7). IEEE.
- [6] Sahoo, J. P., Prakash, A. J., Plawiak, P., & Samantray, S. (2022). Real-time hand gesture recognition using fine-tuned convolutional neural network. *Sensors*, 22(3), 706.
- [7] Varshney, P. K., Kumar, G., Kumar, S., Thakur, B., Saini, P., & Mahajan, V. (2023). Real Time Sign Language Recognition.
- [8] Sen, A., Mishra, T. K., & Dash, R. (2022). Design of Human Machine Interface through vision-based low-cost Hand Gesture Recognition system based on deep CNN. *arXiv preprint arXiv:2207.03112*.
- [9] Howal, A., Golapkar, A., Khan, Y., Bokade, S., Varma, S., & Vyawahare, M. V. (2023, January). Sign Language Finger-Spelling Recognition System Using Deep Convolutional Neural Network. In *2023 5th Biennial International Conference on Nascent Technologies in Engineering (ICNTE)* (pp. 1-6). IEEE.
- [10] Sharma, S., & Saxena, V. P. (2023). Hybrid Sign Language Learning Approach Using Multi-scale Hierarchical Deep Convolutional Neural Network (MDCnn). In *Sentiment Analysis and Deep Learning: Proceedings of ICSADL 2022* (pp. 663-677). Singapore: Springer Nature Singapore.
- [11] Kiran Babu, T. S., & Challa, M. (2022). Recognition of American Sign Language using Deep Learning. *Specialusis Ugdymas*, 1(43), 4069-4075.
- [12] Pranav, P., & Katarya, R. (2022, December). Optimal Sign language recognition employing multi-layer CNN. In *2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)* (pp. 288-293). IEEE.
- [13] Hameed, H., Usman, M., Tahir, A., Ahmad, K., Hussain, A., Imran, M. A., & Abbasi, Q. H. (2022). Recognizing British Sign Language Using Deep Learning: A Contactless and Privacy-Preserving Approach. *IEEE Transactions on Computational Social Systems*.
- [14] Al-Qaisy, N. E., Al-Kaseem, B. R., & Al-Dunainawi, Y. (2023, January). AI-Based Portable Gesture Recognition System for Hearing Impaired People Using Wearable Sensors. In *2023 15th International Conference on Developments in eSystems Engineering (DeSE)* (pp. 33-38). IEEE.
- [15] Lee, M., & Bae, J. (2022). Real-time gesture recognition in the view of repeating characteristics of sign languages. *IEEE Transactions on Industrial Informatics*, 18(12), 8818-8828.
- [16] Padmaja, N., Raja, B. N. S., & Kumar, B. P. (2022). Real time sign language detection system using deep learning techniques. *Journal of Pharmaceutical Negative Results*, 1052-1059.
- [17] Zahid, H., Syed, S. A., Rashid, M., Hussain, S., Umer, A., Waheed, A., ... & Mansoor, N. (2023). A Computer Vision-Based System for Recognition and Classification of Urdu Sign Language Dataset for Differently Aabled People Using Artificial Intelligence. *Mobile Information Systems*, 2023.
- [18] Jayanthi, P., Bhama, P. R., Swetha, K., & Subash, S. A. (2022). Real time static and dynamic sign language recognition using deep learning. *Journal of Scientific & Industrial Research*, 81(11), 1186-1194.
- [19] Tyagi, A., & Bansal, S. (2022). Sign language recognition using hand mark analysis for vision-based system (HMASL). In *Emergent Converging Technologies and Biomedical Systems: Select Proceedings of ETBS 2021* (pp. 431-445). Singapore: Springer Singapore.
- [20] Kandukuri, V., Gundedi, S. R., Kamble, V., & Satpute, V. (2023, April). Deaf and Mute Sign Language Translator on Static Alphabets Gestures using MobileNet. In *2023 2nd International Conference on Paradigm Shifts in Communications Embedded Systems, Machine Learning and Signal Processing (PCEMS)* (pp. 1-6). IEEE.
- [21] Sharma, S., & Singh, S. (2022). Recognition of Indian sign language (ISL) using deep learning model. *Wireless personal communications*, 1-22.
- [22] Omarov, B., Narynov, S., Zhumanov, Z., Gumar, A., & Khassanova, M. (2022). A Skeleton-based Approach for Campus Violence Detection. *Computers, Materials & Continua*, 72(1).
- [23] Furtado, S. L., De Oliveira, J. C., & Shirmohammadi, S. (2023, May). Interactive and Markerless Visual Recognition of Brazilian Sign Language Alphabet. In *2023 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)* (pp. 01-06). IEEE.
- [24] Altayeva, A., Omarov, B., Jeong, H. C., & Cho, Y. I. (2016). Multi-step face recognition for improving face detection and recognition rate.
- [25] Kasapbaşı, A., Elbushra, A. E. A., Omar, A. H., & Yilmaz, A. (2022). DeepASLR: A CNN based human computer interface for American Sign

- Language recognition for hearing-impaired individuals. *Computer Methods and Programs in Biomedicine Update*, 2, 100048.
- [26] Omarov, B., Omarov, B., Shekerbekova, S., Gusmanova, F., Oshanova, N., Sarbasova, A., ... & Sultan, D. (2019). Applying face recognition in video surveillance security systems. In *Software Technology: Methods and Tools: 51st International Conference, TOOLS 2019, Innopolis, Russia, October 15–17, 2019, Proceedings 51* (pp. 271-280). Springer International Publishing.
- [27] Gangrade, J., & Bharti, J. (2023). Vision-based hand gesture recognition for Indian sign language using convolution neural network. *IETE Journal of Research*, 69(2), 723-732.
- [28] Tursynova, A., & Omarov, B. (2021, November). 3D U-Net for brain stroke lesion segmentation on ISLES 2018 dataset. In *2021 16th International Conference on Electronics Computer and Computation (ICECCO)* (pp. 1-4). IEEE.
- [29] Gupta, R., Chaudhary, S., Vedant, A., Choudhury, N. P., & Ladwani, V. (2022). Gesture Detection Using Accelerometer and Gyroscope. In *Emerging Research in Computing, Information, Communication and Applications: Proceedings of ERCICA 2022* (pp. 99-116). Singapore: Springer Nature Singapore.
- [30] Das, S., Biswas, S. K., & Purkayastha, B. (2023). A deep sign language recognition system for Indian sign language. *Neural Computing and Applications*, 35(2), 1469-1481.
- [31] Tazalli, T., Aunshu, Z. A., Liya, S. S., Hossain, M., Mehjabeen, Z., Ahmed, M. S., & Hossain, M. I. (2022, December). Computer vision-based Bengali sign language to text generation. In *2022 IEEE 5th International Conference on Image Processing Applications and Systems (IPAS)* (pp. 1-6). IEEE.
- [32] Khurana, S., Sreemathy, R., Turuk, M., & Jagdale, J. (2023, January). Comparative Study and Performance Analysis of Deep Neural Networks for Sign Language Recognition using Transfer Learning. In *2023 Third International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)* (pp. 1-8). IEEE.
- [33] Sahoo, J. P., Sahoo, S. P., Ari, S., & Patra, S. K. (2022). RBI-2RCNN: Residual block intensity feature using a two-stage residual convolutional neural network for static hand gesture recognition. *Signal, Image and Video Processing*, 16(8), 2019-2027.
- [34] Kapuscinski, T. (2022, June). Handshape Recognition in an Educational Game for Finger Alphabet Practicing. In *International Conference on Intelligent Tutoring Systems* (pp. 75-87). Cham: Springer International Publishing.
- [35] Sultan, D., Omarov, B., Kozhamkulova, Z., Kazbekova, G., Alimzhanova, L., Dautbayeva, A., ... & Abdrakhmanov, R. (2023). A Review of Machine Learning Techniques in Cyberbullying Detection. *Computers, Materials & Continua*, 74(3).
- [36] Mirza, M. S., Munaf, S. M., Azim, F., Ali, S., & Khan, S. J. (2022). Vision-based Pakistani sign language recognition using bag-of-words and support vector machines. *Scientific Reports*, 12(1), 21325.
- [37] Nandi, U., Ghorai, A., Singh, M. M., Changdar, C., Bhakta, S., & Kumar Pal, R. (2023). Indian sign language alphabet recognition system using CNN with diffGrad optimizer and stochastic pooling. *Multimedia Tools and Applications*, 82(7), 9627-9648.
- [38] Pranav, & Katarya, R. (2022). A Systematic Study of Sign Language Recognition Systems Employing Machine Learning Algorithms. In *Distributed Computing and Optimization Techniques: Select Proceedings of ICDCOT 2021* (pp. 111-120). Singapore: Springer Nature Singapore.
- [39] Aggarwal, D., Ahirwar, S., Srivastava, S., Verma, S., & Goel, Y. (2023, March). Sign Language Prediction using Machine Learning Techniques: A Review. In *2023 Second International Conference on Electronics and Renewable Systems (ICEARS)* (pp. 1296-1300). IEEE.
- [40] Singh, S. K., & Chaturvedi, A. (2022, December). Applying Machine Learning for American Sign Language Recognition: A Brief Survey. In *International Conference on Communication and Intelligent Systems* (pp. 297-309). Singapore: Springer Nature Singapore.
- [41] Hussain, M. J., Shaor, A., Alshuibany, S. A., Ghadi, Y. Y., al Shloul, T., Jalal, A., & Park, J. (2022). Intelligent sign language recognition system for E-learning context.
- [42] Robert, E. J., & Duraisamy, H. J. (2023). A review on computational methods based automated sign language recognition system for hearing and speech impaired community. *Concurrency and Computation: Practice and Experience*, 35(9), e7653.
- [43] Amin, M. S., Rizvi, S. T. H., Mazzei, A., & Anselma, L. (2023). Assistive Data Glove for Isolated Static Postures Recognition in American Sign Language Using Neural Network. *Electronics*, 12(8), 1904.
- [44] Hasanov, J., Alishzade, N., Nazimzade, A., Dadashzade, S., & Tahirov, T. (2023). Development of a hybrid word recognition system and dataset for the Azerbaijani Sign Language dactyl alphabet. *Speech Communication*, 102960.

# Non-contact Respiratory Rate Monitoring Based on the Principal Component Analysis

Hoda El Boussaki<sup>1</sup>, Rachid Latif<sup>2</sup>, Amine Saddik<sup>3</sup>, Zakaria El Khadiri<sup>4</sup>, Hicham El Boujaoui<sup>5</sup>  
Laboratory of Systems Engineering and Information Technology LISTI, ENSA,  
Ibn Zohr University, Agadir, Morocco<sup>1,2,3,4,5</sup>  
Faculty of Applied Sciences, Ibn Zohr University, Ait Melloul, Morocco<sup>3</sup>

**Abstract**—Assessing respiratory rate is a critical determinant of one's health status. The proposed approach relies on principal component analysis (PCA) for the continuous monitoring of breathing rate using an RGB camera. This method employs remote plethysmography, a video-based technique enabling contactless tracking of blood volume fluctuations by detecting variations in pixel intensity on the skin. These pixels encompass the red, blue, and green channels, whose values, post-PCA dimensionality reduction, encode the signal containing vital information about the breathing rate. To assess the method's performance, it was tested on a group of seven volunteers, including individuals of both genders. The results reveal a Mean Absolute Deviation of 0.714 BPM and a Root Mean Square Error of 2.035 BPM when comparing the experimental measurements to the actual readings.

**Keywords**—RGB; breathing rate; non-contact; principal component analysis; plethysmography

## I. INTRODUCTION

The respiratory rate is a vital indicator for the driver's current health state. It furnishes insights into clinical deterioration, offers predictive capabilities for cardiac arrest, and aids in the diagnosis of severe pneumonia. It exhibits sensitivity to various pathological conditions like cardiac events as well as stressors, including emotional stress, cognitive load, heat and cold [1]. Alterations and deviations in respiratory rate (RR) are not solely linked to respiratory disorders but also serve as a reliable indicator that a patient is facing challenges in maintaining homeostasis. Respiratory rate acts as an early and highly effective indicator of physiological conditions like hypoxia (insufficient cellular oxygen levels), hypercapnia (elevated carbon dioxide levels in the blood), as well as metabolic and respiratory acidosis. An adult's respiratory rate ranges between 12 and 20 breaths per minute [2]. At this specific respiratory rate, the elimination of carbon dioxide from the lungs matches the body's production of it. However, breathing rates that fall below 12 or exceed 20 may indicate a disturbance in the typical breathing patterns. According to recent findings, an adult who exhibits a respiratory rate exceeding 20 breaths per minute is likely to be in an unhealthy state, while an adult with a respiratory rate surpassing 24 breaths per minute is more likely to be in a critically ill condition [3].

The measurement of the respiratory rate is achieved using sensors, employing a technique that doesn't require direct contact. It quantifies the variation in the reflection of green, blue, and red light from the skin's surface, based on the distinction between specular and diffused reflections [4]. Remote plethysmography (rPPG) is a non-contact method widely used.

It primarily comprises three components, a light source, human skin and a video camera. The light source illuminates the human skin, while the camera records the variations in color [5]. C. Massaroni et al. 2019 presented a method for monitoring the breathing pattern with an RGB camera. The changes in the pixels' intensity gives an overview on the variations of the chest's movements. The system has been tested on 12 volunteers. The Bland-Altman analysis revealed a bias of -0.01 breaths per minute, with respiratory rate values ranging from 10 to 43 breaths per minute [6]. Another method is to use thermal imaging as in the work proposed by Y. Takahashi et al. 2021. Their objective was to monitor the respiration of the subject by measuring temperature variations during exhalation and inhalation. To assess the proposed respiratory rate (RR) estimation method, a study was conducted on seven subjects. The results indicated a mean absolute error of 0.66 beats per minute (bpm) [7]. F. Yang et al. 2022 used an infrared thermal camera to estimate the respiratory rate. The nostril area was chosen as the region of interest and the changes in temperature give an indication on the breathing pattern. The absolute error between the estimated RR and the reference RR from all experiments is  $1.47 \pm 1.33$  breaths/min [8]. J. Kempfle et al. 2020 used a depth camera to estimate the respiratory rate. By capturing and monitoring the subtle changes in distance from the user's chest over time. The findings demonstrate that the method can accurately detect the breathing rate with a range of 92% to 97% from a distance of two meters [9]. P. S. Addison et al. 2023 also used a depth camera. The Bland-Altman analysis revealed limits of agreement of -1.42 to 1.36 breaths/min [10]. Z. El khadiri et al. 2023 proposed an efficient hybrid algorithm for non-contact physiological sign monitoring [11].

In our work, we propose an algorithm that monitors the respiratory rate through an RGB camera. The first part focuses on the face detection and the forehead extraction. The technique proposed by [12] was used for face detection and the extraction of the region of interest. Thus, the box blurring filter, the edge Sobel technique for edge detection, and morphological operations were employed. After that, The raw signal is obtained by computing the mean of each individual channel (red, green, and blue). The signal is then filtered to reduce the noise and the principal component analysis is applied to reduce dimensionality. The resulting signal is then filtered with a bandpass filter with cutoff frequencies of 0.5 and 0.1 Hz corresponding to the breathing rate. Finally, the respiratory rate is calculated by multiplying the maximum frequency after converting the signal to the frequency domain by 60. The summary of our contribution is the proposition of an approach for monitoring the respiratory rate using the

principal component analysis (PCA).

The organization of this paper is as follows: Section II provides an overview of recent advancements in contactless respiratory rate monitoring. Section III outlines our methodology. Subsequently, Section IV presents the results obtained from testing the method on diverse subjects. Finally, the conclusion summarizes the findings of this study and offers insights into future perspectives.

## II. RELATED WORK

Different methods exist to estimate the breathing rate. They are divided into two main categories that are non-contact methods and contact based methods. The first contact based method involves manual human counting. The second method utilizes a spirometer, which provides accurate measurements of respiratory parameters but can interfere with natural breathing and is not suitable for continuous RR monitoring. The third contact based approach involves capnometry but it requires contact with specialized equipment, which may not be comfortable for individuals [13]. Several contactless methods exist to monitor the respiratory rate through a camera. It can be thermal camera, a depth sensing camera or an RGB camera.

Observable fluctuations of the temperature in the region of interest (ROI) that is the nostril or the mouth area are generated by the process of inhalation and exhalation. Microelectromechanical sensors are utilized by thermal imaging cameras to generate images based on heat. The human body becomes distinct within the surrounding environment due to its higher heat emissions. P. Jakkaew et al. 2020 proposed a method that uses thermal imaging to monitor the respiratory rate [14]. The method obtained a root mean square error (RMSE) of  $1.82 \pm 0.75$  bpm. P. Jagadev et al. 2019 employed a thermal camera to monitor the temperature variations across the nostrils during the process of respiration [15]. To automate the tracking of the nostrils (region of interest) despite considerable head movement and object occlusion, a computer vision algorithm called "Ensemble of regression trees" is implemented. The algorithm had a precision of 98.76%. The algorithm demonstrated its effectiveness in managing both stationary and unpredictable head movements. A novel Breath Detection Algorithm (BDA) was introduced to differentiate between normal and abnormal breaths in the acquired breathing waveform. This was achieved by employing predefined thresholds, allowing the algorithm to determine the breaths and calculate the breaths per minute (BPM). A. Kwasniewska et al. 2019 used a Super Resolution (SR) Deep Learning (DL) network to generate enhanced thermal image sequences, which are subsequently analyzed. Despite the improved accuracy achieved through the application of SR algorithms, there is still a significant margin of error remaining [16]. C. B. Pereira et al. 2016 also used infrared thermography (IRT) to monitor the breathing rate. The algorithm takes into account not only the temperature variations around the mouth and nostrils but also the movements of both shoulders [17]. The method was tested in different conditions. The first one is normal breathing and the second one is when there is breathing disorders. During the first condition, a mean correlation of 0.98 and a root-mean-square error (RMSE) of 0.28 bpm was achieved. ON the other hand, the second condition reached a mean correlation of 0.95 and an RMSE of 3.45 bpm. Additionally,

this also showcases the ability of IRT (Infrared Thermography) to effectively capture diverse breathing disorders. L. Chen et al. 2020 introduced a novel approach to non-contact breathing rate (BR) monitoring through a collaborative respiratory detection system. The system utilizes face and motion tracking methods simultaneously to achieve accurate monitoring of the breathing rate [18]. The algorithm showcases its remarkable accuracy with a root mean square error of 0.71 bpm and 0.76 bpm, along with a mean correlation of 0.97. M. Hu et al. 2018 used a combination of near-infrared and thermal imaging techniques for the measurements of breathing rate [19]. For tracking the region of interest (ROI) in thermal video, a tracking algorithm based on spatio-temporal context learning was employed.

In addition to the use of thermal imaging of the mouth and the nostrils, another method is the surveillance of the chest's movements. The expansion of the rib cage occurs during breathing as the diaphragm moves inward and outward. Monitoring the chest movements gives an indication on the number of breaths. In this case depth sensor can extract depth information of the chest area. W. Imano et al. 2020 estimated the respiratory rate from the depth value of the chest and the abdomen. The resulting respiratory rate was compared with the respiratory rate acquired using a spirometer. The experimental results demonstrated that the algorithm achieved a maximum error rate of 1.5% in estimating the respiratory rate [20]. A depth-sensing camera system was also assessed for its performance in continuously monitoring respiratory rate without the need for physical contact in the work of M. Mateumateus et al. 2019. The proposed algorithm involves detecting subject movements using optical flow algorithms on an infrared image. It then calculates the most appropriate region of interest (ROI) that can be utilized by the depth camera to capture the respiratory signal. The algorithm's validity was established by comparing it with a thorax plethysmography system, which served as a reference system [21]. M. Martinez et al. 2017 also used a depth camera to monitor the respiratory rate. The method demonstrates accuracy in 85.9% of the segments, which is comparable to the performance obtained from a chest sensor 88.7%. These results indicate that their use of computer vision is sufficiently precise for the given task.

The third type of cameras that can be used to monitor the breathing rate is RGB cameras. C. Romani et al. 2021 used an RGB camera. Their system enables automated tracking of chest movements associated with breathing, extracting the breathing signal through optical flow and RGB analysis methods. It eliminates events unrelated to breathing from the signal and identifies potential apneas. Additionally, it calculates the respiratory rate value every second [23]. H-S. Hwang et al. 2021 proposed a method for respiration measurement utilizing a region-of-interest detector based on machine learning, in addition to a clustering-based technique to estimate respiration pixels. The proposed approach comprises a model for classifying pixels based on their variance to determine if they convey respiration information. Additionally, a method is employed to classify pixels with distinct breathing components by analyzing the symmetry of the respiration signals [24]. It was established that the average error remained within approximately 0.1 breaths per minute (bpm). H. Ernst et al. 2022 used different combinations of RGB color channels using a hemispherical surface grid search method [25]. The grid search process led to the convergence towards the green channel in the baseline



modulation approach. M. Van Gastel et al. 2016 introduced a non-contact camera-based method for respiratory detection that is capable of operating in both visible and dark lighting conditions. The method relies on detecting the color variations of the skin induced by respiration [26].

### III. METHODOLOGY

The pulsation of the heart generates fluctuations in arterial pressure as blood is pumped through the resistance of the vascular system. Due to the elasticity of arteries, their diameter changes in synchrony with these pressure variations. These alterations in blood volume lead to varying light absorption. Photoplethysmography (PPG) leverages this principle to optically measure blood volume changes by capturing reflected or transmitted light from illuminated skin, resulting in a PPG waveform [27]. When the face is captured, each frame consists of an image composed of three channels: red (R), green (G), and blue (B). The results obtained from Photoplethysmography (PPG) indicate that not only can pulsatility be determined, but also phase information regarding the cardiovascular waveform can be deduced from these three channels [28].

#### A. Region of Interest (ROI) Recognition

Fig. 1 Face detection and ROI extraction. represents the face detection and the extraction of the region of interest algorithm proposed by H. El boussaki et al. 2023 [12].

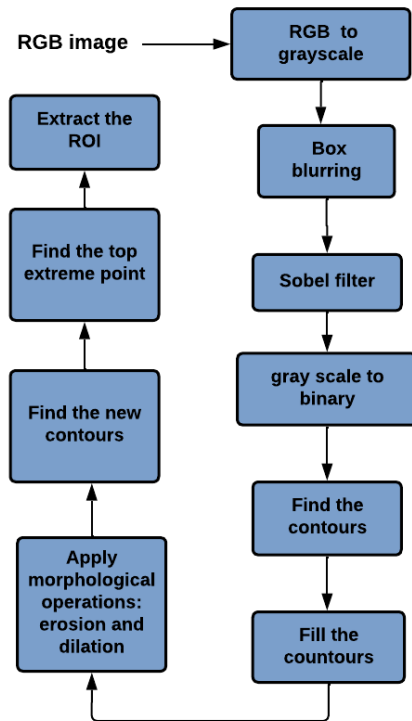


Fig. 1. Face detection and ROI extraction.

The process begins by converting the RGB image to grayscale and then applying a box blurring filter before using the Sobel filter. In the next step, the resulting image is transformed into a binary image, enabling contour detection. Once

the contours are identified, the third step involves filling the interior of the contours with white and applying morphological operations. Finally, the last step involves locating new contours to determine the top extreme point, representing the top of the head. Fig. 1 Face detection and ROI extraction. represented the diagram of the method proposed by [12]. After the top extreme point is detected a value is subtracted from the x coordinate of the top point, and another value is added to the y coordinate. This adjustment enables us to obtain a Region of Interest that starts slightly below the top of the head, precisely where the forehead is located.

#### B. Signal Extraction

The raw signal is obtained from the image by employing a function that computes the average of the pixels in each channel (red, green, and blue). The averages of these channels are then combined to form the signal. The RGB components within the region of interest (ROI) are spatially averaged across all pixels, resulting in an RGB component for each frame. These averaged RGB components form the raw signals. As new frames are processed, their values are added to the signal. At this point, the signal reflects the variations in pixel values from one frame to another.

#### C. Signal Filtering

The signal underwent additional denoising using a band-pass filter. This filter had a lower cutoff frequency of 0.1 Hz and a higher cutoff frequency of 0.5 Hz. When multiple channels are employed, the signal's dimensionality is commonly decreased by combining the channels in a linear manner. The Principal Component Analysis (PCA) is a well-known for its ability to reduce dimensionality. The PCA is applied on the filtered signal. It generates three linearly uncorrelated components, which are obtained by combining the three RGB signals in a linear fashion. The PCA is then a linear technique for reducing dimensionality, transforming a set of correlated features from a high-dimensional space into a sequence of uncorrelated features in a lower-dimensional space [29]. These uncorrelated features, known as principal components, are produced as a result [30]. It is a linear transformation that is orthogonal, indicating that all the principal components are perpendicular to one another. It reshapes the data in a manner where the first component endeavors to account for the highest amount of variance present in the original data. PCA aids in identifying the most prominent feature within a dataset, simplifies the representation of data in 2D and 3D plots, and facilitates the discovery of a sequence of linear combinations of variables [31]. The central aspect of PCA is dimensionality reduction, which involves reducing the number of dimensions within a given dataset. When the data exhibits a clear linear trend and directed points, applying PCA allows for straightforward reduction of the dimensional data into a lower-dimensional representation. The objective of PCA is to identify a new matrix that represents the principal components. This matrix captures the essential information and structure of the original data represented by X, an  $m \times n$  matrix. Y is an  $m \times n$  matrix that is connected through a linear transformation represented by P and is a re-representation of X as shown in Eq. (1) [32].

$$Y = PX \quad (1)$$

Where P represents the matrix that transforms Y into X.

Then the covariance of X is computed. The covariance quantifies the strength of the linear relationship between two variables. A high value indicates a strong positive relationship, while a low value suggests a weak or no relationship. It is represented in Eq. (2) [33].

$$C_x = \frac{1}{n-1}XX^T \quad (2)$$

Where  $C_x$  is the square symmetric matrix, the diagonal terms of  $C_x$  are the variance of particular measurement types and the off-diagonal terms of  $C_x$  are the covariance between measurement types.

$C_x$  encompasses the correlations among all potential pairs of measurements, with the correlation values indicating the presence of noise and redundancy in our measurements. The objective is to acquire a matrix Y in such a way that the covariance matrix exhibits the highest variance. PCA operates under the assumption that P is an orthonormal matrix. Additionally, it assumes that the directions with the highest variances correspond to the most significant signals, making them the principal directions.  $C_y$  in terms of our variable P is represented in Eq. (3).

$$C_y = \frac{1}{n-1}YY^T = \frac{1}{n-1}PAP^T = \frac{1}{n-1}PXX^TP^T \quad (3)$$

$C_y$  is a symmetric matrix, whose eigenvalues are arranged on the principal diagonal of the matrix A in descendent order, and the eigenvectors constitute the columns of the matrix P. The principal components of X are the eigenvectors of  $XX^T$  or the rows of P. Performing PCA on a dataset X involves subtracting the mean of each measurement type and then computing the eigenvectors of the matrix  $XX^T$  [33].

Fig. 2Signal filtering. represents the filtering algorithm and summarizes the previous steps. The signal goes through a denoising filter and a normalization, then a bandpass filter and the principal component analysis and a moving average filter.

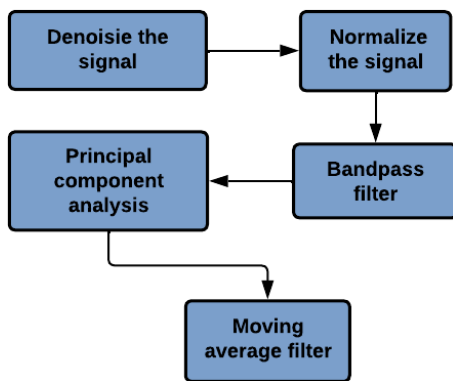


Fig. 2. Signal filtering.

Typically, the PCA technique uses tabular information and data, the rows stand in for the observations you want to incorporate and embed in a place with less dimensional space, while the columns correspond to the features for which you are looking for a reduced approximation. The principal components are

generated by performing the singular value decomposition after the algorithm has calculated the covariance matrix in minute detail. Since smaller data sets are easier to examine, explore, visualize, and make analyzing data much easier and faster for machine learning algorithms without extraneous variables to process, the trick in dimensionality reduction is to trade a little accuracy for simplicity. For more convenience, the following pseudo-code illustrates the prominent steps for the Principal Component Analysis (PCA) technique:

---

**Algorithm 1** Principal Component Analysis - PCA

---

Consider Z to be a data array of size nxm  
Center and standardize the data array  
 $Y \leftarrow \frac{Z - \mu}{\sigma}$  while  $\mu$  is the mean, and  $\sigma$  is the standard deviation  
Calculate the covariance matrix of Y  
 $Y \leftarrow Y^TY$   
Calculate the eigenvectors and eigenvalues of  $Y^TY$   
Sort the eigenvalues from largest to smallest  
 $\lambda_1 > \lambda_2 > \dots > \lambda_p$   
Sort the eigenvectors in the matrix P accordingly  
 $Y^* \leftarrow YP$   
Calculate the proportion of variance explained for each feature  
Add features with the highest explained proportion of variation until it reaches a certain threshold

---

**D. Respiratory Rate Estimation**

A discrete Fourier transform is used to convert the resulting signal to the frequency domain [34]. The maximum of the frequency index is extracted as the frequency corresponding to the breathing. The respiratory rate is calculated with Eq. (4) [35]. The algorithm takes a sequence of images as input and identifies a Region of Interest (RoI). For each pixel within the region of interest, it constructs a trajectory in the time domain. This trajectory represents the pixel values across the entire sequence.

$$BPM = Max * 60 \quad (4)$$

Where Max is the maximum frequency  
Fig. 3Respiratory rate estimation. represents the respiratory rate calculation algorithm. The discrete Fourier transform is applied to the signal. Then, if there is enough data and that means that the signal is large enough, the power spectrum with the highest magnitude is extracted. The value is used in Eq. (4) to calculate the respiratory rate.

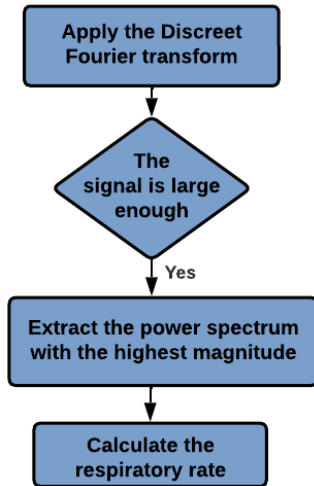


Fig. 3. Respiratory rate estimation.

#### IV. EXPERIMENTAL RESULTS

A dataset was gathered, comprising seven volunteers, including four females and three males with a mean age of 37.8 years and an age range of 18 to 58 years. All participants provided informed consent for the test experiment. The volunteers was positioned at a distance of approximately 1 meter from the camera and instructed to blink and breathe naturally. This work was implemented on an Intel i7-1165G7 desktop using its camera. The respiratory rate was calculated through the proposed algorithm and compared with the values obtained by counting the number of breaths for one minute. In this paper, the performance of the respiratory rate measurement method is evaluated using the following indicators: the Mean Absolute Deviation (MAD) [36] and Root Mean Square Error (RMSE) [37]. The first metric represents the average absolute error between the estimated respiratory rate and the reference estimation. It provides insights into the accuracy of the measured respiratory rate compared to the desired respiratory rate and is calculated with Eq. (5) [38].

$$MAD = \frac{1}{n} \sum |RR_{rppg}^i - RR^i| \quad (5)$$

Where  $RR_{rppg}$  is the respiratory rate estimated through an RGB camera and  $RR$  is the respiratory rate calculated manually. The second metric is calculated with equation 6 [38].

$$RMSE = \sqrt{\frac{\sum (RR_{rppg}^i - RR^i)^2}{n}} \quad (6)$$

Table I Respiratory Rate Obtained in Different Subjects represents the breathing rate obtained from seven volunteers that consists of three males and four females. The respiratory rate obtained using the method proposed was than compared with the respiratory rate acquired by counting the number of breaths per minute.

TABLE I. RESPIRATORY RATE OBTAINED IN DIFFERENT SUBJECTS

Subjects	Gender	Respiratory rate estimated (BPM)	Reference (BPM)
Subject 1	M	15	16
Subject 2	F	17	18
Subject 3	F	20	20
Subject 4	F	21	20
Subject 5	M	17	17
Subject 6	M	22	17
Subject 7	F	23	22

The evaluation metrics to assess the deviation of the measurement results from the reference breathing rate were employed to verify the accuracy of the measurement results. The calculated Mean Absolute Deviation (MAD) is 0.714 bpm and the calculated Root Mean Square Error is 2.035 bpm. Y. Takahashi et al. 2021 used a thermal camera and their method was tested on seven subjects with a mean absolute error of 0.66 beats per minute. Yang et al. 2022 used an infrared thermal camera and the absolute error is  $1.47 \pm 1.33$  breaths per minute. P. Jakkaew et al. 2020 proposed a method that uses a thermal camera and obtained a root mean square error of  $1.82 \pm 0.75$  bpm. C. B. Pereira et al. 2016 also used infrared thermography and achieved a root-mean-square error (RMSE) of 3.45 breaths per minute. C. Romano et al. 2021 used an RGB camera and obtained a bias of  $-0.03 \pm 1.38$  bpm and  $-0.02 \pm 1.92$  bpm in the Bland Altman analysis. H-S. Hwang and E. C. Lee 2021 proposed a method that was tested and evaluated using data from 14 men and women in a real-world environment using convolutional neural networks. During this evaluation, it was found that the correlation coefficient between the contactless signal and the reference signal being 0.93 on average indicates a strong positive linear relationship between the two signals. This suggests that our method's performance was quite accurate compared to others cited before. Our method gives a nearly same or an even higher performance compared to the use of other methods. However, the use of convolutional neural networks gives better performances.

#### V. CONCLUSION

This paper introduces a non-contact heart rate monitoring algorithm designed to measure the respiratory rate of the driver. The proposed method shows a Mean Absolute Deviation of 0.714 BPM and a Root Mean Square Error of 2.035 BPM. The approach consists of detecting the top extreme point of the head through image filtering, contour finding and morphological operations. When the top extreme point is detected, it is easy to determine the region of interest as it is located under the top extreme point. The signal is extracted from the changes in the pixels' intensity. Then, the signal is filtered and the principal component analysis is applied. Future works consist of evaluating the algorithm's processing time, improving it through parallel programming and implementing it in various embedded architectures.

#### ACKNOWLEDGMENT

We owe a debt of gratitude to the Ministry of National Education, Vocational Training, Higher Education and Scientific Research (MENF-PERSRS) and the National Center for Scientific and Technical Research of Morocco (CNRST) for their financial support (grant number: 27UIZ2022) and for the project Cov/2020/109.

#### REFERENCES

- [1] A. Nicolò, C. Massaroni, E. Schena, and M. Sacchetti, "The importance of respiratory rate monitoring: From healthcare to sport and exercise," *Sensors*, 20(21), 6396, 2020.
- [2] A. Rowden, "What is a normal respiratory rate based on your age?" January 2023.
- [3] Respiratory rate. (n.d.). Physiopedia. [https://www.physio-pedia.com/Respiratory\\_Rate](https://www.physio-pedia.com/Respiratory_Rate)
- [4] A. Bella, R. Latif, A. Saddik, and L. Jamad, "Review and Evaluation of Heart Rate Monitoring Based Vital Signs, A case Study: Covid-19 Pandemic," 6th IEEE Congress on Information Science and Technology (CiSt), 2020.
- [5] A. Bella, R. Latif, A. Saddik, and F. Z. Guerrouj, "Monitoring of Physiological Signs and Their Impact on The Covid-19 Pandemic: Review. E3S Web of Conferences, 229, 01030, 2021.
- [6] C. Massaroni, D. Lo Presti, D. Formica, S. Silvestri, and E. Schena, "Non-Contact monitoring of breathing pattern and respiratory rate via RGB signal measurement," *Sensors*, 19(12), 2758, 2019.
- [7] Y. Takahashi, Y. Gu, T. Nakada, R. Abe, and T. Nakaguchi, "Estimation of Respiratory Rate from Thermography Using Respiratory Likelihood Index," *Sensors*, 21(13), 4406, 2021.
- [8] F. Yang, S. He, S. Sadanand, A. Yusuf, and M. Bolic, "Contactless measurement of vital signs using thermal and RGB cameras: A study of COVID 19-Related Health Monitoring," *Sensors*, 22(2), 627, 2022.
- [9] J. Kempfle, and K. Van Laerhoven, "Towards breathing as a sensing modality in Depth-Based Activity recognition," *Sensors*, 20(14), 3884, 2020.
- [10] P. S. Addison, A. Antunes, D. Montgomery, P. Smit, and U. R. Borg, "Robust Non-Contact Monitoring of Respiratory Rate using a Depth Camera," *Journal of Clinical Monitoring and Computing*, 2023.
- [11] Z. El Khadiri, R. Latif, and A. Saddik, "An efficient hybrid algorithm for non-contact physiological sign monitoring using plethysmography wave analysis," *Computer Methods in Biomechanics and Biomedical Engineering. Imaging & Visualization*, 1–17, 2023.
- [12] H. El Boussaki, R. Latif, and A. Saddik, "Video-based Heart Rate Estimation using Embedded Architectures," *International Journal of Advanced Computer Science and Applications*, 14(5), 2023.
- [13] M. C. T. Manullang, Y. Lin, S. Flaxman, and N. Chou, "Implementation of Thermal Camera for Non-Contact Physiological Measurement: A Systematic Review," *Sensors*, 21(23), 7777, 2021.
- [14] P. Jakkaw and T. Onoye, "Non-Contact Respiration Monitoring and Body Movements Detection for Sleep Using Thermal Imaging," *Sensors (Basel)*. 2020 Nov 5;20(21):6307.
- [15] P. Jagadev, and L. I. Giri, "Non-contact monitoring of human respiration using infrared thermography and machine learning," *Infrared Physics & Technology*, 104, 103117, 2020.
- [16] A. Kwasniewska, M. Szankin, J. Ruminski, and M. Kaczmarek, "Evaluating Accuracy of Respiratory Rate Estimation from Super Resolved Thermal Imagery," 2029.
- [17] C. B. Pereira, X. Yu, M. Czaplík, V. Blazek, B. Venema, and S. Leonhardt, "Estimation of breathing rate in thermal imaging videos: a pilot study on healthy human subjects," *Journal of Clinical Monitoring and Computing*, 31(6), 1241–1254, 2016.
- [18] L. Chen, M. Hu, N. Liu, G. Zhai, and S. X. Yang, "Collaborative use of RGB and thermal imaging for remote breathing rate measurement under realistic conditions," *Infrared Physics & Technology*, 111, 103504, 2020.
- [19] M. Hu, G. Zhai, B. Yang, Y. Fan, H. Duan, W. Zhu, and M. Yang, "Combination of near-infrared and thermal imaging techniques for the remote and simultaneous measurements of breathing and heart rates under sleep situation," *PLOS ONE*, 13(1), e0190466, 2018.
- [20] W. Imano, K. Kameyama, M. Hollingdal, J. C. Refsgaard, K. S. Larsen, C. S. R. Topp, S. H. Kronborg, J. D. Gade, and B. Dinesen, "Non-Contact respiratory measurement using a depth camera for elderly people," *Sensors*, 20(23), 6901, 2020.
- [21] M. Mateu-Mateus, F. Guede-Fernandez, M. A. Garcia-Gonzalez, J. Ramos-Castro, and M. Fernandez-Chimeno, "Non-Contact Infrared-Depth Camera-Based method for respiratory rhythm measurement while driving," *IEEE Access*, 7, 152522–152532, 2019.
- [22] M. Martínez, and R. Stiefelhagen, "Breathing Rate Monitoring during Sleep from a Depth Camera under Real-Life Conditions," 2017.
- [23] C. Romano, E. Schena, S. Silvestri, and C. Massaroni, "Non-Contact respiratory monitoring using an RGB camera for Real-World applications," *Sensors*, 21(15), 5126, 2021.
- [24] H-S. Hwang, and E. C. Lee, "Non-Contact Respiration Measurement method based on RGB camera using 1D convolutional neural networks," *Sensors*, 21(10), 3456, 2021.
- [25] H. Ernst, H. Malberg and M. Schmidt, "Non-contact Measurement of Respiration Rate with Camera-based Photoplethysmography During Rest and Mental Stress," 2022 Computing in Cardiology (CinC), Tampere, Finland, pp. 1–4, 2022.
- [26] M. Van Gastel, S. Stuijk, and G. De Haan, "Robust respiration detection from remote photoplethysmography," *Biomedical Optics Express*, 7(12), 4941, 2016.
- [27] Z. El Khadiri, R. Latif, and A. Saddik, "Breathing pattern assessment through the empirical mode decomposition and the empirical Wavelet transform algorithms," In *Lecture notes on data engineering and communications technologies*, pp. 262–271, 2023.
- [28] W. Verkrusse, L. O. Svaasand, and J. M. Nelson, "Remote plethysmographic imaging using ambient light," *Optics Express*, 16(26), 21434, 2008.
- [29] Z. El Khadiri, R. Latif, and A. Saddik, "Remote heart rate measurement using plethysmographic wave analysis," In *Lecture notes in networks and systems*, pp. 254–267, 2023.
- [30] K. Parte, "Dimensionality reduction: principal component analysis," *Medium*, December 2021.
- [31] V. Lendave, "Detecting Orientation of Objects in Image using PCA and OpenCV," *Analytics India Magazine*, 2021.
- [32] E. De Benedetto, A. Bottiglieri, S. Pisa, and M. Cavagnaro, "Cardiorespiratory frequency monitoring using the principal component analysis technique on UWB Radar Signal," *International Journal of Antennas and Propagation*, 1–6, 2017.
- [33] J. Shlens, "A tutorial on principal component analysis," 2003.
- [34] S. Mejhoubi, R. Latif, W. Jenkal, A. Saddik, and A. El Ouardi, "Hardware Architecture for Adaptive Dual Threshold Filter and Discrete Wavelet Transform based ECG Signal Denoising," *International Journal of Advanced Computer Science and Applications*, 12(11), 2021.
- [35] H. El Boussaki, R. Latif, and A. Saddik, "A review on Video-Based Heart Rate, Respiratory Rate and Blood Pressure Estimation," In *Lecture notes in networks and systems*, pp. 129–140, 2023. H. El Boussaki, R. Latif, and A. Saddik, "A review on Video-Based Heart Rate, Respiratory Rate and Blood Pressure Estimation," In *Lecture notes in networks and systems*, pp. 129–140, 2023.
- [36] O. E. B'charri, R. Latif, K. Elmansouri, A. Abenaou, and W. Jenkal, "ECG signal performance de-noising assessment based on threshold tuning of dual-tree wavelet transform," *Biomedical Engineering Online*, vol. 16, no. 1, Feb. 2017.
- [37] W. Jenkal, R. Latif, A. Toumanari, A. Dliou, O. E. B'charri, and F. M. R. Maoulainine, "An efficient algorithm of ECG signal denoising using the adaptive dual threshold filter and the discrete wavelet transform," *Biocybernetics and Biomedical Engineering*, vol. 36, no. 3, pp. 499–508, Jan. 2016.
- [38] B. Zhang, H. Li, L. Xu, L. Qi, Y.-D. Yao, and S. E. Greenwald, "Noncontact heart rate measurement using a webcam, based on joint blind source separation and a skin reflection model: for a wide range of imaging conditions," *Journal of Sensors*, vol. 2021, pp. 1–18, Jul. 2021.

# An Improved Genetic Algorithm with Chromosome Replacement and Rescheduling for Task Offloading

Hui Fu, Guangyuan Li, Fang Han\*, Bo Wang

Faculty of Engineering, Huanghe Science and Technology College, Zhengzhou, China 450006

**Abstract**—End-Edge-Cloud Computing (EECC) has been applied in many fields, due to the increased popularity of smart devices. But the cooperation of end devices, edge and cloud resources is still challenge for improving service quality and resource efficiency in EECC. In this paper, we focus on the task offloading to address the challenge. We formulate the offloading problem as mixed integer nonlinear programming, and solve it by Genetic Algorithm (GA). In the GA-based offloading algorithm, each chromosome is the code of a offloading solution, and the evolution is to iteratively search the global best solution. To improve the performance of GA-based task offloading, we integrate two improvement schemes into the algorithm, which are the chromosome replacement and the task rescheduling, respectively. The chromosome replacement is to replace the chromosome of every individual by its better offspring after every crossing, which substitutes the selection operator for population evolution. The task rescheduling is rescheduling each rejected task to available resources, given offloading solution from every chromosome. Extensive experiments are conducted, and results show that our proposed algorithm can improve upto 32% user satisfaction, upto 12% resource efficiency, and upto 35.3% processing efficiency, compared with nine classical and up-to-date algorithms.

**Keywords**—Genetic algorithm; task offloading; task scheduling; edge computing; cloud computing

## I. INTRODUCTION

Smart devices, such as smartphones, Internet of Things (IoT) devices, drones, and so on, have become ubiquitous in our life and their number continues to grow rapidly [1], as communications and information technology advance and our quality of life improves. Unfortunately, due to their small physical space, most devices have limited resource capacity and battery life. As a result, devices frequently lack the processing power required by user requests, especially for complex applications like facial recognition and intelligent driving, which become more and more common.

To address the above issue, several works make use of cloud computing, which provides “infinite” computing resources, to extend the processing capacity of devices [2], [3]. However, cloud computing has poor network performance because it typically provides services via a Wide Area Network (WAN), such as the Internet. To address this issue, edge computing brings a few computing resources (edge servers) close to devices to provide low latency services [4], [5], [6]. Combining advantages of end devices, edge servers and cloud, end-edge-cloud computing (EECC) has attracted much attention from both academia and industry, as it can effectively and efficiently provide various services to end users [7], [8], [9].

In EECC environments, it is challenge to efficiently utilize the collaboration of devices, edge servers and cloud. To address the challenge, several works have designed task offloading or scheduling algorithms for EECC, to improve service performance or/and resource efficiency. The task offloading is to decide the computing node for each task’s processing. Existing works have made some assumptions to simplify the offloading problem for EECC, which limits their application scope. For example, some works ignored the heterogeneity between edge and cloud resources, which can lead to resource inefficiency [10], [11]. Some works didn’t exploit the resource capacity of end devices, even though many modern devices are equipped with a wealth of hardware resources, and thus wasted zero-delay local resources for task processing.

There are mainly two categories algorithms used for task offloading, heuristics and meta-heuristics. Heuristics exploit some local optimum search strategies tailored to the specific problem. Heuristics generally have rapid solving processes but limited performances. In contrast, meta-heuristics are general problem solvers. Meta-heuristics apply both local searches and global searches, inspired by natural and social rules. Usually, compared with heuristics, meta-heuristics can achieve better performance, but cost more time.

Therefore, in this paper, we exploit hybridization of heuristics and meta-heuristics, to exploit their complementary strengths for the task offloading in EECC, considering the resource heterogeneity. Even though some works have proposed hybrid heuristic offloading algorithms, most of them simply perform two or more algorithms sequentially, leading to a poor performance of hybridization. Specifically, we use genetic algorithm (GA) due to its representativeness and extensive application. GA has powerful global search ability, but slow convergence sometimes. To make up for this shortcoming, we propose to use chromosome replacement instead of selection operator for GA. To improve the performance of offloading solutions decoded from chromosomes, we reschedule failed tasks by a heuristic algorithm. The contributions of this paper can be summarized as follows:

- The task offloading problem of EECC is formulated into a mixed integer nonlinear programming problem (MINLP), with deadline constraints. The optimization objectives are maximizing the finished task number and the overall resource utilization, which are commonly used for quantifying user satisfaction and resource efficiency, respectively.
- A task offloading algorithm is proposed based on GA and first fit heuristic scheduling (FF). The proposed algorithm uses an integer coding approach for mapping

\*Corresponding authors.

between task offloading solutions and chromosomes. GA is employed for searching the global best solution. To improve the quality of task offloading solutions, FF is used to reschedule failed tasks to available resources in EECC, for every offloading solution. To speed up the convergence of GA, the selection operator is replaced by a replacement operator that replaces every chromosome with its better offspring produced by crossover.

- Extensive simulated experiments are conducted for evaluating the performance of our proposed algorithm. Simulation parameters are set referring to related works. Experiment results verify that our proposed algorithm can finish more tasks than nine of classical and up-to-date offloading algorithms. The efficiencies of the replacement and the task rescheduling are also verified by the results.

The content below is organised as follows. Section II formulates the task offloading problem in EECC. Section III illustrates our proposed offloading algorithm. Section IV evaluates the performance of the proposed algorithm. Section V presents the works related to task offloading for EECC. Section VI concludes this paper.

## II. PROBLEM STATEMENT

### A. Resource and Task Model

In this paper, we consider the EECC system consisting of  $D$  end devices,  $E$  edge servers (ES), and  $V$  cloud servers (CS). We use  $s_i, 1 \leq i \leq D + E + V$  to represent these computing nodes, where devices include  $s_i, 1 \leq i \leq D$ , ES are  $s_i, D+1 \leq i \leq D + E$ , and CS are  $s_i, D + E + 1 \leq i \leq D + E + V$ . For computing node  $s_i$ , there are  $n_i$  computing cores each with  $g_i$  capacity. The network connection between two computing nodes, say  $s_i$  and  $s_j$ , is represented by constants  $b_{i,j}$  which is its data transfer rate. If there is no connection between  $s_i$  and  $s_j$ ,  $b_{i,j} = 0$ . For each computing node, there is no data transmission delay within it, i.e.,  $b_{i,i} = +\infty, 1 \leq i \leq D + E + V$ .

$T$  tasks ( $t_k, 1 \leq k \leq T$ ) are launched by  $D$  devices. Binary constants  $o_{i,k}, 1 \leq i \leq D, 1 \leq k \leq T$  are used to indicate the ownerships of these tasks, where  $o_{i,k} = 1$  means  $t_k$  is launched by  $s_i$ , and  $o_{i,k} = 0$  means not. Task  $t_k$  has  $c_k$  computing size, i.e., it requires  $c_k$  computing resource for its processing. The input data amount of  $t_k$  is  $a_k$ . In this paper, we ignore the transmission delay of the output data, because the output data amount usually is very small [12]. The deadline of  $t_k$  is  $d_k$ , which means  $t_k$  must be finished before  $d_k$ . For every task, if its deadline constraint cannot be satisfied, it will be rejected, because there will be no profit for processing the task.

A task offloading solution is the mapping/assignments of tasks to computing cores for their processing, which can be represented by a set of binary variables  $x_{i,j,k}$ , as shown in Eq. 1.  $x_{i,j,k}$  is 1 if  $t_k$  is assigned to  $j$ th core in computing node  $s_i$ , and 0 otherwise. In this paper, we consider the resource granularity as computing core instead of computing node, because considering fine granularity of resources helps to improve the resource efficiency [13].

$$x_{i,j,k} = \begin{cases} 1, & \text{if } t_k \text{ is assigned to } j\text{th core in } s_i \\ 0, & \text{else} \end{cases}, \quad (1)$$

$$1 \leq i \leq D + E + V, 1 \leq j \leq n_i, 1 \leq k \leq T.$$

For each task, it can be only assigned to one core for its processing. In this paper, we don't consider to use the redundant execution for the performance improvement due to its huge resource costs. Thus, Eq. 2 holds.

$$\sum_{i=1}^{D+E+V} \sum_{j=1}^{n_i} x_{i,j,k} \leq 1, 1 \leq k \leq T. \quad (2)$$

And when  $t_k$  is accepted and processed by a computing core,  $\sum_{i=1}^{D+E+V} \sum_{j=1}^{n_i} x_{i,j,k} = 1$ . When  $t_k$  is rejected,  $\sum_{i=1}^{D+E+V} \sum_{j=1}^{n_i} x_{i,j,k} = 0$ . Then, the number of accepted tasks can be achieved by Eq. 3.

$$N = \sum_{k=1}^T \sum_{i=1}^{D+E+V} \sum_{j=1}^{n_i} x_{i,j,k}. \quad (3)$$

### B. Task Processing Model

For a task assigned to a core, its computing can be started only when its input data transfer finishes and the core is available. Then Eq. (4) must be satisfied, where  $ft_k^A$  and  $st_k$  are respectively the completion time of data transfer and the start time of computing for  $t_k$ .

$$ft_k^A \leq st_k, 1 \leq k \leq T. \quad (4)$$

When  $t_k$  is assigned to  $j$ th core in the computing node  $s_i$ , its computing consumes  $c_k/g_i$  time. Its input data is transferred from its device ( $s_{i'}$  where  $o_{i',k} = 1$ ) to  $s_i$ . Then the data transfer rate is  $\sum_{i'}^D (o_{i',k} \cdot b_{i',i})$ , and the transfer time is  $a_k / \sum_{i'}^D (o_{i',k} \cdot b_{i',i})$ . Therefore, for each task, the start time and the finish time of data transfer and computing satisfy constraints is Eq. (5) and (6), where  $st_k^A$  and  $ft_k$  represent the start time of  $t_k$ 's input data transfer and the finish time of its computing, respectively. Noticing that Eq. (5)–(6) also hold for rejected tasks, as both sides of inequality operators are 0 for these tasks. Then, the deadline constraints can be formulated as Eq. (7).

$$st_k^A + \frac{a_k}{\sum_{i=1}^{D+E+V} \sum_{j=1}^{n_i} (x_{i,j,k} \cdot \sum_{i'}^D (o_{i',k} \cdot b_{i',i}))} \leq ft_k^A, \quad (5)$$

$$1 \leq k \leq T.$$

$$st_k + \frac{c_k}{\sum_{i=1}^{D+E+V} \sum_{j=1}^{n_i} (x_{i,j,k} \cdot g_i)} \leq ft_k, 1 \leq k \leq T. \quad (6)$$

$$ft_k \leq d_k, 1 \leq k \leq T. \quad (7)$$

When multiple tasks are assigned to one computing core, they cannot be computed simultaneously. There are two cases

for the computing order of two tasks ( $t_k$  and  $t_{k'}$ ) in every computing core. If  $t_k$  is computing before  $t_{k'}$ ,  $ft_k \leq st_{k'}$ , and otherwise,  $ft_{k'} \leq st_k$ . Therefore, Eq. (8) formulates the exclusiveness of tasks' computing in every computing node.

$$x_{i,j,k} \cdot x_{i,j,k'} \cdot (st_{k'} - ft_k) \cdot (st_k - ft_{k'}) \leq 0, \quad (8)$$

$$1 \leq k, k' \leq T.$$

For each computing core, its occupied time is the latest finish time of tasks assigned to it, which is  $\tau_{i,j} = \max_{k=1}^T (x_{i,j,k} \cdot ft_k)$  for  $j$ th core of  $s_i$ . The occupied time of a computing node is the maximal occupied time of its cores, which is  $\tau_i = \max_{j=1}^{n_i} \max_{k=1}^T (x_{i,j,k} \cdot ft_k)$  for  $s_i$ . Thus, the amount of occupied computing resources is  $\tau_i \cdot n_i \cdot g_i$  on  $s_i$ . While, the resources effectively use for computing tasks are  $R = \sum_{k=1}^T \sum_{i=1}^{D+E+V} \sum_{j=1}^{n_i} (x_{i,j,k} \cdot c_k)$  in overall system. Then, the overall computing resource can be calculated by Eq. (9).

$$U = \frac{R}{\sum_{i=1}^{D+E+V} (\tau_i \cdot n_i \cdot g_i)}. \quad (9)$$

### C. Task Offloading Problem Model

Now, based on above formulations, the task offloading problem can be modelled as follows:

$$\text{Maximizing } N + U, \quad (10)$$

subject to,

$$\text{Eq. (1)–(9)}. \quad (11)$$

The objective is to maximize the number of accepted tasks plus the overall computing resource utilization. Because the total number of tasks is fixed in EECC system, the maximization of accepted task number is identical to maximizing the accepted ratio which is one commonly used metric for quantifying user satisfaction, service level agreement, and quality of service. As resource utilization is not greater than one, user satisfaction maximization is the major optimization objective in the model. Noticing that  $U$  is nonlinear and non-convex, and decision variables including binary ( $x_{i,j,k}$ ) and continuous variables ( $st_k^A$ ,  $ft_k^A$ ,  $st_k$ , and  $ft_k$ ), the task offloading problem belongs to mixed integer non-linear programming (MINLP), which is hard to be solved exactly. In fact, the task offloading problem has been proved NP-hard [14]. Therefore, in the next section, we propose a hybrid heuristic algorithm for efficiently solving the offloading problem with polynomial time.

### III. IMPROVED GENETIC ALGORITHM FOR OFFLOADING

In this section, we design an improved genetic algorithm with chromosome replacement and task rescheduling method, GRRS, to solve the task offloading problem presented in the previous section, which is outlined in Algorithm 1.

At first, we design a solution representation method (or encoding/decoding approach) to create the map between task offloading solutions and chromosomes used for search in GA-based algorithms. For GA inspired by Charles Darwin's theory of evolution, the population consists of multiple individuals and is evolved by changing these individuals' chromosomes each with multiple genes. In the solution representation

---

**Algorithm 1** GRRS: The genetic offloading algorithm with replacement and rescheduling

---

**Input:** The information of tasks, and EECC resources;

**Output:** A task offloading strategy;

- 1: Initializing chromosomes of individuals randomly;
  - 2: Evaluating fitness of every individual using Algorithm 2;
  - 3: Initializing the best chromosome ( $bc$ ) as one with the best fitness in all individuals;
  - 4: **while** the terminal condition is not reached **do**
  - 5:   **for** each individual ( $Y$ ) **do**
  - 6:     Crossing  $Y$  with another individual which is randomly selected, with a certain probability, and producing two offspring, i.e., new chromosomes;
  - 7:     Evaluating the fitnesses of two offspring;
  - 8:     Replacing  $Y$ 's chromosome with the better offspring;
  - 9:     **if**  $Y$ 's chromosome has better fitness than  $bc$  **then**
  - 10:       Updating  $bc$  as  $Y$ 's chromosome;
  - 11:     **end if**
  - 12:     mutating  $Y$  with a certain probability;
  - 13:     Evaluating fitness of  $Y$ ;
  - 14:     Updating  $bc$  as done in lines 9–11.
  - 15:   **end for**
  - 16: **end while**
  - 17: **return** the offloading strategy decoded from  $bc$  by Algorithm 2;
- 

method, there is a one-to-one relationship between genes and tasks in EECC. The value of each gene is integer, which identifies the computing core where the corresponding task is assigned. Thus, the possible value of a gene is 1 to the number of cores which can be used for processing the corresponding task. Then, we get the assignments of all tasks from a chromosome or an individual.

For example, considering an EECC system consisting two devices, two ES, and two CS, where each node has one computing core and each device launches one task. Then, the number of genes is 2, corresponding to these two tasks. The possible value in each dimension is 1 to 5, respectively representing the cores of the device, two ES, and two CS for the corresponding task.

A fitness function is needed to evaluate how goodness of every chromosome/individual. We use the optimization objective (10),  $N + U$ , as the fitness function of GRRS. The fitness evaluation of every chromosome is given in Algorithm 2. Given a chromosome, it can be easily to achieve a task assignment solution by the solution representation (lines 3–5 in Algorithm 2). To achieve a complete offloading solution, the computing order of tasks assigned to every core needs to be decided. In this paper, we use the most simple algorithm, First Fit (FF), for order decisions, and will study more efficient algorithms to improve the performance of GRRS. With FF order, we can calculate the finish time of each task one-by-one (line 7 in Algorithm 2). If the finish time fits deadline constraint for a task, it is accepted, and otherwise, rejected (lines 8–12 in Algorithm 2). After all tasks are decided to be accepted or rejected, GRRS tries to reschedule every rejected task using FF, to improve the overall user satisfaction (lines 14–22 in Algorithm 2). Now, we achieve a task offloading solution from a chromosome. The accepted task number and the overall resource utilization can be easily calculated based on the offloading solution, and the fitness is achieved for the chromosome.

---

**Algorithm 2** Decoding a chromosome into a task offloading with rescheduling

---

**Input:** A chromosome;

**Output:** A task offloading solution,  $\Omega$ , and the fitness;

```
1:  $\Omega \leftarrow \phi$ ; /*the set including assignments of accepted tasks to
   computing cores*/
2:  $\Phi \leftarrow \phi$ ; /*the set including rejected tasks*/
3: for each gene of the chromosome do
4:   Per-assigning the corresponding task into the computing core
   identified by the gene value;
5: end for
6: for each task  $t$  do
7:   Calculating its finish time in the scheme of first fit scheduling
   on the core ( $c$ ) to which it is per-assigned;
8:   if the finish time is earlier than the deadline then
9:      $\Omega \leftarrow \Omega \cup \{< t, c >\}$ ; /*deciding to assign the task to the
   core*/
10:  else
11:     $\Phi \leftarrow \Phi \cup \{t\}$ ; /*the deadline being violated*/
12:  end if
13: end for
14: for each task  $t \in \Phi$  /*rescheduling rejected tasks*/ do
15:   for each core  $c$  do
16:     Calculating its finish time as line 7;
17:     if the finish time is earlier than the deadline then
18:        $\Omega \leftarrow \Omega \cup \{< t, c >\}$ ; /*rescheduling  $t$  to  $c$ */
19:       break; /*rescheduling another rejected task*/
20:     end if
21:   end for
22: end for
23: Calculating overall resource utilization  $U$  by Eq. 9;
24: return  $\Omega$  and  $|\Omega| + U$ ; /*the accepted task number  $N = |\Omega|$ */
```

---

Based on the solution representation and the fitness evaluation, GRRS exploits the main idea of GA to iteratively search for the optimal solution for task offloading, as shown in Algorithm 1. First, GRRS initializes the population, i.e., randomly sets the value of every gene for every individual's chromosome, as done by standard GA. Then, GRRS evaluates the fitness for every initialized chromosome, and sets the best chromosome as the chromosome with the best fitness. After these initialization steps, GRRS uses some operators to evolve individuals by updating their chromosomes (lines 4–16 in Algorithm 1). The evolution procedure is as follows.

First, for every individual, GRRS uses crossover operator to create new chromosomes/offspring. GRRS randomly selects another individual, and performs the crossover operator on them, with a certain possibility (the crossover possibility). To ensure the individual diversity for large-scale offloading problems, GRRS exploits the uniform crossover operator, which swaps values of two chromosomes in every gene location with a certain probability. After crossing an individual, two new chromosomes are produced. For the individual, GRRS evaluates the fitness of its two offspring, and replaces its chromosome by the offspring with better fitness than another one. By such replacement, GRRS can increase the diversity by retaining new produced chromosomes, and speed the convergence rate by transmitting good genes of the better offspring to the next generation. If a new offspring has better fitness than the best chromosome, the best one is updated as the offspring.

To further enhance exploration ability, GRRS applies the uniform mutation operator on each individual, to increase

the diversity by creating new genes. The uniform mutation operator is to change each gene with the mutation possibility for an individual. After mutating an individual, if the new chromosome has better fitness than the best one, the best one is updated as the new one.

GRRS repeats the above evolution procedure until the terminal condition is reached. There are two approaches for the set of terminal condition. One is setting the maximal number of iterations, and another is setting the most times that the fitness of the best chromosome has no (significant) change. After the evolution procedure, GRRS decodes the best chromosome into the task offloading solution, and return it as the global best solution.

In this paper, we focus on the improvement of GA by chromosome replacement and task rescheduling. Undoubtedly, the crossover and mutation operators as well as the parameters have impact on the performance of GA. These opportunities will be studied on our future works.

#### IV. PERFORMANCE EVALUATION

In this section, we conduct extensive simulated experiments to verify the efficiency of GRRS by comparing with several classical and up-to-date offloading algorithms. The experiment environment is illustrated in Section IV-A, and the results are discussed in Section IV-B.

##### A. Experiment Environment

In simulated EECC systems, where the simulation parameters are set referring to [18], [21], [23], [15] and reality, there are ten devices, five ES, and ten types of CS. The core number of devices, ES, and CS are set randomly in ranges of [2,8], [4,32], and [1,8], respectively. The computing capacities of each core in every device, ES, and CS are respectively set as [1.8,2.5]GHz, [1.8,3.0]GHz, and [1.8,3.0]GHz, randomly. The network transfer rate between a device and an ES/CS is in the range of [80,120]/[10,20] Mbps. There are 1000 tasks generated, and the device for lunching each task is randomly allocated. The computing resource required by a task is in the range of [0.5, 1.2]GHz, and the input data amount is [1.5, 6]MB. The deadline of every task is set between one and five seconds.

The algorithms used for the performance comparison with GRRS to confirm the performance include FF, FFD, EDF, RAND, GA, GAR, PSO, PSOM, and GAPSO.

- First Fit (FF) iteratively schedules the first task to the first computing node meeting its requirements.
- First Fit Decreasing (FFD) iteratively schedules the task requiring maximal amount of computing resources to the first computing node meeting its requirements.
- Earliest Deadline First (EDF) iteratively schedules the task with the earliest deadline to the first computing node meeting its requirements.
- Random method (RAND) randomly generates a population as done by GRRS, and provides the solution corresponding to the best individual.



- GA [15], [16] uses the uniform crossover, the uniform mutation, and the roulette wheel selection operators for the population evolution.
- GA with replacement (GAR) [17] is same to GRRS without the rescheduling.
- Particle Swarm Optimization (PSO) [18] uses the idea of particle movement. PSO initializes a population with multiple particle (individual), and iteratively moves each particle toward its personal best position and the global best position for the particle position updates (the population evolution).
- PSO with mutation operator (PSOM) [19] added a mutation operator on each particle at the end of each iteration.
- GAPSO [20] first initializes a population, and then sequentially performs GA and PSO on the population evolution.

The performance of above algorithms are evaluated as following, and all performance metrics are better when their values are greater.

- User satisfaction is the experience of users, which has great influence on the profit and the reputation of service providers. In this paper, we use the number of tasks with deadline met, i.e., the accepted task number ( $N$ ), for the quantification.
- Resource efficiency is the workload processed by a unit of resources, which determines the cost-performance of service provision. The metrics used for measuring the resource efficiency are the computing resource utilization ( $U$ , the completed computing size per unit of computing resource) and the data processing efficiency (the processed data amount per unit of computing resource,  $\sum_{k=1}^T \sum_{i=1}^{D+E+V} \sum_{j=1}^{n_i} (x_{i,j,k} \cdot a_k) / \sum_{i=1}^{D+E+V} (\tau_i \cdot n_i \cdot g_i)$ ).
- Processing efficiency is the processing speed of a computing system, which is quantified by the completed computing size and the processed data amount by a time unit ( $R / \max_i \{\tau_i\}$  and  $\sum_{k=1}^T \sum_{i=1}^{D+E+V} \sum_{j=1}^{n_i} (x_{i,j,k} \cdot a_k) / \max_i \{\tau_i\}$ ).

The experiment process are as follows. We first generate a EECC system, and then sequentially measure various performance metrics for all of comparison algorithms and GRRS. For each measured value for every algorithm and every metric, we normalize it by dividing it into that of FF, to focus on the relative performance between different algorithms. These previous experiment steps are repeated more than 100 times, and we report the average value for each metric in the follows. Noticing that, in each measurement, there is a new EECC system generated randomly. Thus, the statistical information of every algorithm in each metric is meaningless without the normalization. GRRS has statistically significant difference with other algorithms in every performance metric.

Besides comparing the performance of GRRS with other offloading algorithms, we verify the efficiency of the task rescheduling, by comparing the performance between

GA/PSO/GAPSO/GAR with and without task rescheduling. The results are presented and discussed in section IV-B4.

## B. Experiment Results

1) *User Satisfaction*: Fig. 1 gives the relative number of accepted tasks when applying different task offloading methods, on average. As shown in the figure, GRRS achieves the most accepted tasks, which completes 7.98%–32% more tasks than other algorithms. This verifies that GRRS performs good on the optimization of the user satisfaction. The main reasons are as follows.

For heuristics, FF, FFD, and EDF, the priority order of resources used for task processing is devices, ES, and CS. This can complete more task in low network latency but scarce resources of devices and ES. As shown in experiment results, GRRS accepts 12%–29.4% and 32%–33.7% less tasks than these heuristic algorithms at devices and ES, respectively, as shown in Fig. 2 and Fig. 3. But this can result in some tasks with loose deadline assigned to devices or ES at first. This leads to insufficient resources for processing subsequent tasks with tight deadline, and thus can drastically decrease the number of tasks processed by CS and reduce the overall user satisfaction. As shown in Fig. 4, meta-heuristics process more than 100% more tasks than heuristics by the cloud. As shown in Fig. 2 - 4, GRRS processes not the most tasks in one tier of devices, edges, and cloud. But GRRS has the best overall satisfaction, as shown in Fig. 1. This phenomenon verifies the powerful global search ability of GRRS.

Compared with other meta-heuristic algorithms (GA, GAR, PSO, PSOM, GAPSO), GRRS achieve better performance in optimizing the user satisfaction, as shown in Fig. 1. The main advantages of GRRS is the replacement replacing the selection operator for GA and the rescheduling for improving the quality of the solution corresponded to an individual. GAR can complete more tasks than GA, which verifies that the replacement improves the evolution effectiveness by substituting the selection operator. The improvement of the rescheduling strategy on the task offloading will be illustrated in Section IV-B4.

In addition, we can see that some meta-heuristics (GA, PSO, PSOM, and GAPSO) has poorer performance than heuristics, even though they are designed for pursuing the global best and heuristics are aiming at the local best, in such a large-scale offloading problem. This inspires us that meta-heuristics should be carefully designed for a good performance.

2) *Resource Efficiency*: Fig. 5 and Fig. 6 show resource utilization and data processing efficiency achieved by different offloading algorithms. From the figure, we can see that heuristics achieve higher resource utilization and higher data processing efficiency than meta-heuristics. This is mainly because that heuristics process tasks using scarce local and edge resources at first. This can provide a good performance for accepted tasks, and a high computing resource efficiency, because there is no or a low latency for data transfer. But with prioritization of device and edge resources, much less tasks can be completed by the cloud resources, leading to a low user satisfaction as illustrated above. Contrary to heuristics, meta-heuristics try to find the global best solution, which can schedule every task to

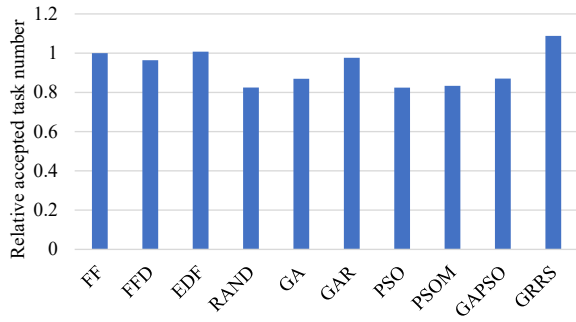


Fig. 1. The relative accepted task number achieved by various task offloading methods in overall.

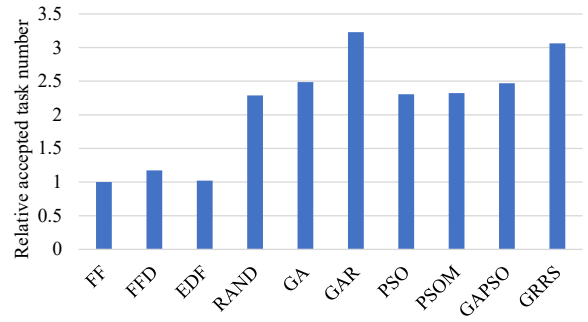


Fig. 4. The relative accepted task number achieved by various task offloading methods in the cloud.

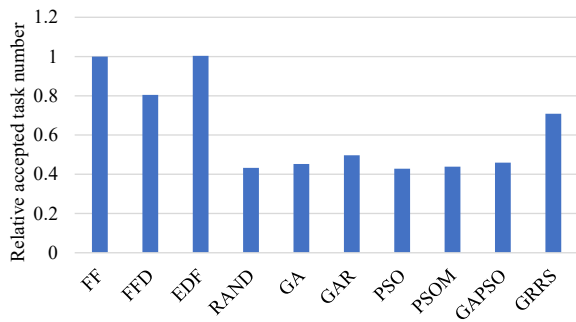


Fig. 2. The relative accepted task number achieved by various task offloading methods in device tier.

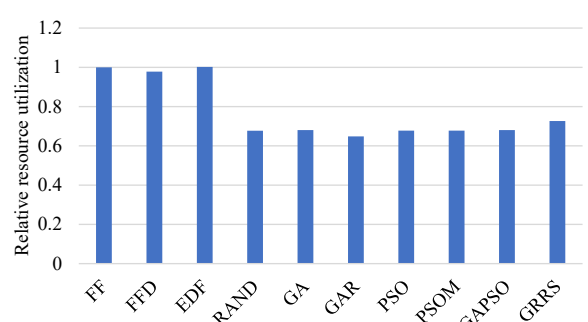


Fig. 5. The relative resource utilization achieved by various task offloading methods.

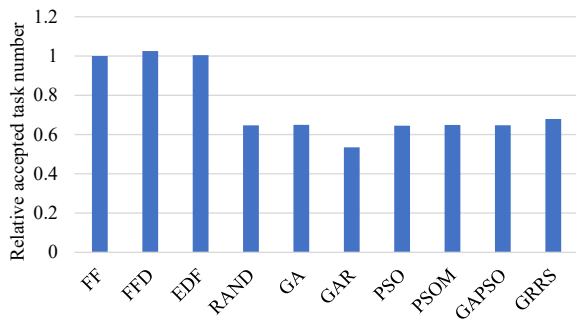


Fig. 3. The relative accepted task number achieved by various task offloading methods in edge tier.

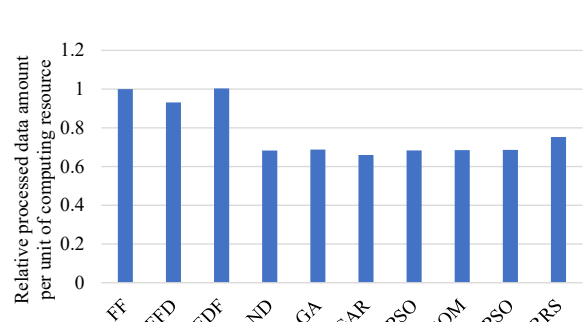


Fig. 6. The relative processed data amount per unit of computing resources achieved by various task offloading methods.

any resource at first. This provides opportunities for processing more tasks in overall. Therefore, GRRS improves the user satisfaction by sacrificing the resource efficiency, compared with heuristics. It is worth it because the user satisfaction usually decides the income and reputation of service providers.

In all of these meta-heuristics, GRRS has the highest resource utilization and data processing efficiency, which are 6.74%–12% and 9.5%–14.1% higher than that of other meta-heuristics, respectively, as shown in Fig. 5 and 6. This demonstrates that GRRS performs good at the optimization of both the user satisfaction and the resource efficiency, and further confirms the high effectiveness of GRRS.

3) *Processing Efficiency*: Fig. 7 and Fig. 8 present the processing rates or efficiencies in computing and data processing in EECC when applying various task offloading methods. In a distributed system, the processing rate reflects the parallelism, and thus the throughput, which is one of the most used metrics quantifying overall performance. As shown in these two figures, we can see that GRRS has the highest processing rates, which are 7.6%–31.5% and 11.4%–35.3% higher than other methods in the computing and data processing, respectively. This illustrates that GRRS achieves good processing efficiency for EECC systems.

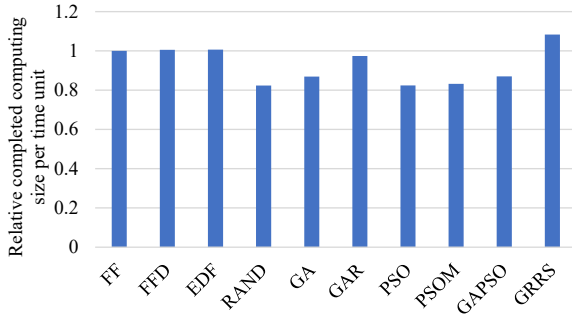


Fig. 7. The relative computing efficiency by various task offloading methods.

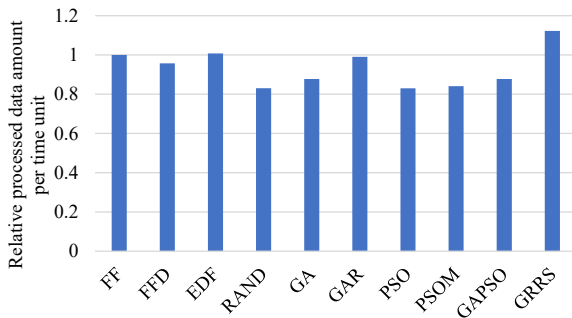


Fig. 8. The relative data processing efficiency achieved by various task offloading methods.

4) *Performance of Improvements*: One of our improvement schemes for meta-heuristics is task rescheduling, which can be applied by any meta-heuristic. In this section, we test the performance of rescheduling in the improvement of user satisfaction, resource efficiency, and processing efficiency. The experiment results are shown in Fig. 9-13, where  $x_{RS}$  means  $x$  improved with the rescheduling. From these figures, we can see that rescheduling can improve above 11% performance in every metric for meta-heuristic algorithms on task offloading. This verifies the high efficiency of rescheduling in improving the performance of meta-heuristic-based offloading algorithms. The main reason why rescheduling can improve the performance of meta-heuristics is that meta-heuristics make decision of task assignment without considering the load balance between computing cores. This leads to some cores are overloaded while some others are underloaded, giving a opportunity for performance improvement by rescheduling.

## V. RELATED WORKS

As the development of IoT, EECC has been applied to various fields for improving the performance of various data processing applications. To improve service quality and resource efficiency in EECC environments, several works focused on addressing the task offloading problem.

Sang et al. [21] proposed a heuristic offloading algorithm to improve the cooperativeness of EECC resources. They used cloud resources for processing offloaded tasks at first, and rescheduled some tasks from the cloud to ES and devices to

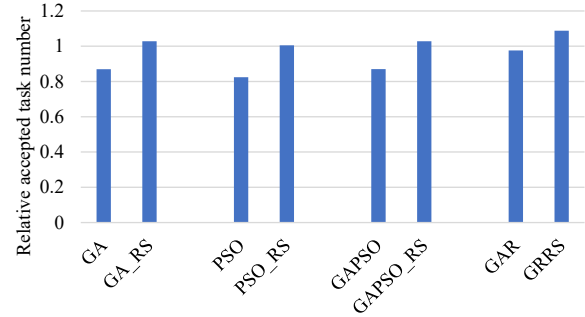


Fig. 9. The accepted task number improved by rescheduling.

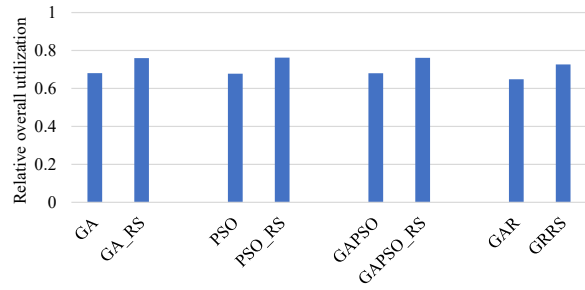


Fig. 10. The resource utilization improved by rescheduling.

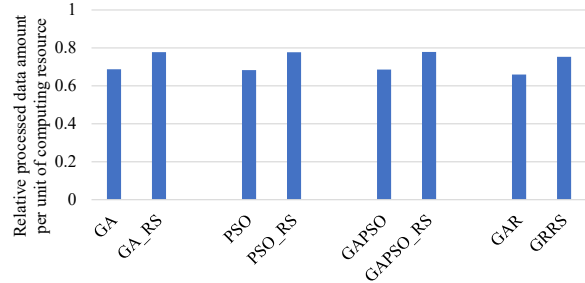


Fig. 11. The data processing efficiency improved by rescheduling.

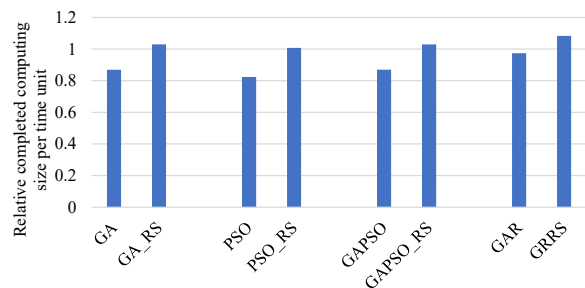


Fig. 12. The overall computing rate improved by rescheduling.

improve overall performance. This can improve overall user satisfaction, but negatively affect the overall performance of task processing. Wang et al. [22] presented two offloading al-

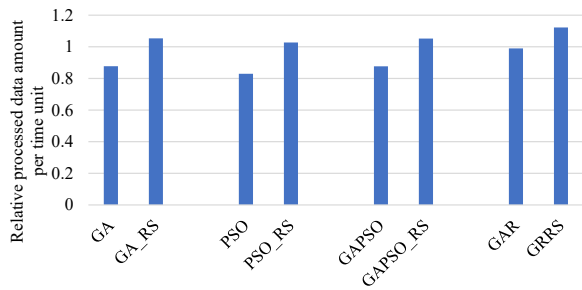


Fig. 13. The overall data processing rate improved by rescheduling.

gorithms, named as FRFOA and LBOA, respectively. FRFOA was to offload a task to an ES such that the response time is minimum every time. LBOA iteratively assigned a task to an ES which can satisfy requirements of the most tasks. These heuristic-based algorithms generally consume few resources but have limited performance, because they only exploit local search strategies.

Therefore, some works used meta-heuristics to pursue the global best offloading solution. Both Wang et al. [18] and Gao et al. [23] applied PSO with same solution representation method to this paper. In addition, to improve exploration ability, Gao et al. [23] used Lévy Flight movement pattern for updating particle positions. Wang et al. [15] used GA for optimizing user satisfaction and resource efficiency. Chakraborty and Mazumdar [16] employed GA to reduce energy consumption with latency constraints. Bali et al. [24] used NSGA-II to optimize energy and queue delay for offloading data to ES and CS.

To further improve performance, some works considered to exploit complementary advantage of different algorithms, proposed hybrid heuristic algorithms. Hussain and Al-Turjman [17] replaced the chromosome by its better offspring generated by the crossover operator for each individual, which is similar to population evolution behavior of PSO. Nwogbaga et al. [19] performed mutation operator for every individual at the end of each evolutionary iteration for PSO to improve diversity for avoiding premature convergence. Farsi et al. [20] sequentially performed GA and PSO for the population evolution. Zhang et al. [25] presented a dynamic selection mechanism to combine multiple meta-heuristics, which selected offspring generated by these meta-heuristics to be passed on to next generation. All of above works just performed two or more meta-heuristics separately, which leads to a poor performance of combination.

Therefore, in this paper, we exploit a combination approach to integrate the swarm intelligent into the evolutionary algorithm for a better offloading solution on EECC. In addition, we propose to use heuristic rescheduling approach to further improve the solution quality.

## VI. CONCLUSION

In this section, we focus on the task offloading problem for EECC systems. We first formulate the problem into MINLP, which has been proofed as NP-hard. Then, to solve the problem with reasonable time complexity, we design a task offloading algorithm, GRRS, based on GA which is one of the most

representative meta-heuristics and performs well on solving various optimization problems in many fields. To enhance exploration and exploitation of GA, we integrate two improvement scheme into it. One is replacing each individual with its better offspring during the population evolution, to pass on good genes. Another is rescheduling rejected tasks to take full advantage of available EECC resources. Extensive experiments are conducted, and results verify efficiency and effectiveness of GRRS.

## ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions. The research was supported by the key scientific and technological projects of Henan Province (Grant No. 232102211084, 232102210023, 232102210125), the Key Scientific Research Projects of Henan Higher School (Grant No. 22A520033), Zhengzhou Basic Research and Applied Research Project (ZZSZX202107) and China Logistics Society (2022CSLKT3-334).

## REFERENCES

- [1] Cisco Systems, Inc., Cisco Annual Internet Report (2018–2023) White Paper, <https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/white-paper-c11-741490.html>, Mar. 2020.
- [2] X. Jin, W. Hua, Z. Wang and Y. Chen, “A survey of research on computation offloading in mobile cloud computing,” *Wireless Networks*, 2022, 28(4): 1563–1585, May 2022, doi: 10.1007/s11276-022-02920-2.
- [3] C. Bi, J. Li, Q. Feng, C.-C. Lin and W.-C. Su, “Optimal deployment of vehicular cloud computing systems with remote microclouds,” *Wireless Networks*, February 2023, In Press, 13 pages, doi: 10.1007/s11276-023-03268-x.
- [4] W. Shi, J. Cao, Q. Zhang, Y. Li and L. Xu, “Edge Computing: Vision and Challenges,” *IEEE Internet of Things Journal*, 2016, 3(5): 637-646, doi: 10.1109/JIOT.2016.2579198.
- [5] X. Wang, J. Li, Z. Ning, Q. Song, L. Guo, S. Guo, and M. S. Obaidat. “Wireless Powered Mobile Edge Computing Networks: A Survey,” *ACM Computing Surveys*, January 2023, In Press, doi:10.1145/3579992.
- [6] M. Reiss-Mirzaei, M. Ghojaei-Arani, L. Esmaili, “A review on the edge caching mechanisms in the mobile edge computing: A social-aware perspective,” *Internet of Things*, 2023, vol. 22, Article ID: 100690, 22 pages, doi:10.1016/j.iot.2023.100690.
- [7] J. Ren, D. Zhang, S. He, Y. Zhang and T. Li, “A Survey on End-Edge-Cloud Orchestrated Network Computing Paradigms: Transparent Computing, Mobile Edge Computing, Fog Computing, and Cloudlet,” *ACM Computing Surveys*, 2019, vol. 52, no. 6, Article ID: 125, 36 pages, doi: 10.1145/3362031
- [8] T. Wang, Y. Liang, X. Shen, X. Zheng, A. Mahmood, and Q. Z. Sheng, “Edge Computing and Sensor-Cloud: Overview, Solutions, and Directions,” *ACM Computing Surveys*, February 2023, In Press, doi: 10.1145/3582270
- [9] B. Kar, W. Yahya, Y. -D. Lin and A. Ali, “Offloading Using Traditional Optimization and Machine Learning in Federated Cloud-Edge-Fog Systems: A Survey,” *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, pp. 1199-1226, Secondquarter 2023, doi: 10.1109/COMST.2023.3239579.
- [10] M. I. Khaleel, “Efficient job scheduling paradigm based on hybrid sparrow search algorithm and differential evolution optimization for heterogeneous cloud computing platforms,” *Internet of Things*, 2023, vol. 22, Article ID: 100697, 29 pages, doi: 10.1016/j.iot.2023.100697.
- [11] W. Khallouli and J. Huang, “Cluster resource scheduling in cloud computing: literature review and research challenges,” *The Journal of Supercomputing*, vol. 78, pp. 6898–6943, 2022. doi: 10.1007/s11227-021-04138-z.

- [12] W. Lu, Y. Mo, Y. Feng, Y. Gao, N. Zhao, Y. Wu, and A. Nallanathan, "Secure Transmission for Multi-UAV-Assisted Mobile Edge Computing Based on Reinforcement Learning," *IEEE Transactions on Network Science and Engineering*, vol. 10, no. 3, pp. 1270-1282, 1 May-June 2023, doi: 10.1109/TNSE.2022.3185130.
- [13] Z. Liu, C. Chen, J. Li, Y. Cheng, Y. Kou, D. Zhang, "KubFBS: A fine-grained and balance-aware scheduling system for deep learning tasks based on kubernetes," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 11, Article ID: e6836, 16 pages, 2022. doi:10.1002/cpe.6836
- [14] J. Du, and J. Y.-T. Leung, "Complexity of Scheduling Parallel Task Systems," *SIAM Journal on Discrete Mathematics*, vol. 2, no. 4, pp. 473-487, 1989, doi: 10.1137/0402042.
- [15] B. Wang, B. Lv, and Y. Song, "A Hybrid Genetic Algorithm with Integer Coding for Task Offloading in Edge-Cloud Cooperative Computing," *IAENG International Journal of Computer Science*, vol. 49, no. 2, pp. 503-510, 2022.
- [16] S. Chakraborty, K. Mazumdar, "Sustainable task offloading decision using genetic algorithm in sensor mobile edge computing," *Journal of King Saud University - Computer and Information Sciences* vol. 34, no. 4, pp. 1552-1568, 2022, doi: 10.1016/j.jksuci.2022.02.014.
- [17] A. A. Hussain and F. Al-Turjman, "Hybrid Genetic Algorithm for IOMT-Cloud Task Scheduling," *Wireless Communications and Mobile Computing*, vol. 2022, Article No. 6604286, 14 pages, 2022, doi: 10.1155/2022/6604286.
- [18] B. Wang, J. Cheng, J. Cao, C. Wang and W. Huang, "Integer particle swarm optimization based task scheduling for device-edge-cloud cooperative computing to improve SLA satisfaction," *PeerJ Computer Science*, vol. 8, Article ID: e893, 22 pages, 2022. doi:10.7717/peerj-cs.893
- [19] N. E. Nwogbaga, R. Latip, L. S. Affendey, and A. R. Abdul Rahiman, "Attribute reduction based scheduling algorithm with enhanced hybrid genetic algorithm and particle swarm optimization for optimal device selection," *Journal of Cloud Computing*, vol. 11, Article ID: 15, 17 pages, 2022. doi: 10.1186/s13677-022-00288-4.
- [20] A. Farsi, S. Ali Torabi, and M. Mokhtarzadeh, "Integrated surgery scheduling by constraint programming and meta-heuristics," *International Journal of Management Science and Engineering Management*, July 2022, In Press, doi: 10.1080/17509653.2022.2093289.
- [21] Y. Sang, J. Cheng, B. Wang and M. Chen, "A three-stage heuristic task scheduling for optimizing the service level agreement satisfaction in device-edge-cloud cooperative computing," *PeerJ Computer Science*, vol. 8, Article ID: e851, 24 pages, 2022, doi:10.7717/peerj-cs.851
- [22] C. Wang, R. Guo, H. Yu, Y. Hu, C. Liu, C. Deng, "Task offloading in cloud-edge collaboration-based cyber physical machine tool," *Robotics and Computer-Integrated Manufacturing*, vol. 79, Article ID: 102439, 13 pages, 2023, doi: 10.1016/j.rcim.2022.102439.
- [23] T. Gao, Q. Tang, J. Li, Y. Zhang, Y. Li and J. Zhang, "A Particle Swarm Optimization With Lévy Flight for Service Caching and Task Offloading in Edge-Cloud Computing," *IEEE Access*, vol. 10, pp. 76636-76647, 2022, doi: 10.1109/ACCESS.2022.3192846.
- [24] M. S. Bali, K. Gupta, D. Gupta, G. Srivastava, S. Juneja, and A. Nauman, "An effective technique to schedule priority aware tasks to offload data on edge and cloud servers," *Measurement: Sensors*, vol. 26, Article ID: 100670, 9 pages, 2023, doi:10.1016/j.measen.2023.100670.
- [25] J. Zhang, Z. Ning, R. H. Ali, M. Waqas, S. Tu and I. Ahmad, "A Many-objective Ensemble Optimization Algorithm for the Edge Cloud Resource Scheduling Problem," *IEEE Transactions on Mobile Computing*, In Press, January 2023, 18 pages, doi: 10.1109/TMC.2023.3235064.

# An Efficient Convolutional Neural Network Classification Model for Several Sign Language Alphabets

Ahmed Osman Mahmoud, Ibrahim Ziedan, Amr Ahmed Zamel  
Computer and Systems Department, Faculty of Engineering, Zagazig Uni., Zagazig, Egypt

**Abstract**—Although deaf people represent over 5% of the world’s population, according to what the World Health Organization stated in May 2022, they suffer from social and economic marginalization. One way to improve the lives of deaf people is to try to make communication between them and others easier. Sign language, the means through which deaf people can communicate with other people, can benefit from modern techniques in machine learning. In this study, several convolutional neural networks (CNN) models are designed to develop an efficient model, in terms of accuracy and computational time, for the classification of different signs. This research presents a methodology for developing an efficient CNN architecture from scratch to classify multiple sign language alphabets, which has numerous advantages over other contemporary CNN models in terms of prediction time and accuracy. This framework analyses the effect of varying CNN hyper-parameters, such as kernel size, number of layers, and number of filters in each layer, and picks the ideal parameters for CNN model construction. In addition, the suggested CNN architecture operates directly on unprocessed data without the need for preprocessing to generalize it across other datasets. In addition, the capacity of the model to generalize to diverse sign languages is rigorously evaluated using three distinct sign language alphabets and five datasets, namely, Arabic (ArSL), two American English (ASL), Korean (KSL), and the combination of Arabic and American datasets. The proposed CNN architecture (SL-CNN) outperforms state-of-the-art CNN models and traditional machine learning models achieving an accuracy of 100%, 98.47%, 100%, and 99.5% for English, Arabic, Korean, and combined Arabic-English alphabets, respectively. The prediction or inference time of the model is about three milliseconds on average, making it suitable for real-time applications. So, in the future, it is easy to turn this model into a mobile application.

**Keywords**—Convolutional neural network (CNN); sign language; Arabic sign language (ArSL); American sign language (ASL); Korean sign language (KSL); Complexity time

## I. INTRODUCTION

The World Health Organization (WHO) stated in May 2022 that there are more than 360 million deaf people around the world. 80% of those who are deaf live in developing countries and use more than 300 sign languages [1]. Deaf people who suffer from hearing problems have trouble in their daily lives communicating with each other and with other people. Also, they have lower chances of having an adequate level of education.

Sign language is the means of communication between deaf people and other people and consists of hand signs and gestures for spelling letters and words. In the last two decades, many researchers have investigated several machine learning models

for developing several sign language recognition models, e.g., Arabic [2], American [3], Korean [4], Indian, Chinese, and others [5]. The approaches for sign language recognition can be classified into two main categories: sensor-based and vision-based [6]. In a sensor-based, the speaker wears gloves or sensors, and the movement and body orientation are translated into a time series of sensor readings depending on the word or letter sign. Within the same category, several researchers use Microsoft Kinect [7], developed by Microsoft, for sign classification without wearing gloves or sensors. Microsoft Kinect has three optical sensors. It provides three outputs: an RGB image, an infrared (IR) image, or a depth image, and defines up to 25 skeleton joints. Generally, the other hand, in the vision-based approach [8], images are taken by a camera and analyzed to determine the shape of signs intended by the speaker. In this approach, a researchers use depth images and skeleton joints to analyze the kinematic movement of the body or hand to determine the word or alphabet character. In this way, sign classification became easier. On sufficient number of image examples for each sign should be collected to improve classification performance.

Several methods for sign language recognition, utilizing both traditional and deep learning techniques, have been proposed in the literature [5]. Traditional techniques such as Support Vector Machine (SVM) [9], Hidden Markov Model (HMM) [10], and Random Forest (RF) [11] have all been tried by many researchers for classifying sign language alphabet recognition, but they have all yielded unsatisfactory results. On the other hand, recent studies have shown that the CNN model is one of the most commonly used models in sign language recognition [6]. Surveys such as those conducted by Rastgou et al. have shown that many models have been suggested by various researchers for sign language recognition with the help of deep learning techniques [5].

### A. Motivation

Many researchers have tried traditional machine learning methods for classifying sign language alphabet recognition models [12], but they have not provided satisfactory results. Recently, the CNN model has been widely utilized in sign language recognition, but it has not offered an efficient CNN architecture, which is considerably more challenging due to CNN’s numerous hyper-parameters. Moreover, the prediction time, which is the time to predict a single sign and a critical factor in practice, is usually ignored. In addition, although CNN model hyper-parameters have a large effect on performance and accuracy [13], they are not fully investigated. All

these issues may affect model generalization for newly unseen data. These are the primary motivations for us to study the effect of selecting hyper-parameters in CNN and how we can generalize the CNN architecture across several datasets. The primary objective of this study is to provide a realistic, straightforward, and efficient CNN architecture. In addition, it can work on several sign language alphabets.

## B. Contribution

In this paper, several CNN models are designed and analyzed to develop an optimal model for sign language recognition without any preprocessing.

- This research provided a method for selecting an optimal CNN model by analyzing the results of various hyper-parameters such as kernel size, number of layers, and number of filters in each layer.
- In the beginning, this approach investigated how changing the layer and filter numbers impacted accuracy.
- Then, the model with the highest accuracy and the fastest prediction time is chosen.
- In the end, we incorporate the impact of kernel size and the number of hidden layers to pick the best candidate.
- This approach operates directly on unprocessed data to generalize it across several datasets.
- The proposed model is examined on four different datasets, Arabic, American 2018, American 2012, and Korean alphabets, to investigate its robustness against data variation.
- In addition, the model is also tested on a combined dataset of Arabic and American alphabets.

This paper is organized as follows: Section II provides a literature review on Arabic, English, and Korean sign language detection. In Section III, a detailed design of the proposed CNN model is presented, including the selection of optimal hyper-parameters for Arabic and English sign languages. The proposed model is presented in Section IV. Several experiments are then conducted in Section V to evaluate and compare the proposed model to other state-of-the-art classification models for three sign language alphabets. Finally, conclusions and ideas for extending the current work are drawn in Section VI.

## II. LITERATURE REVIEW

Sign language differs from country to country. For example, there are Arabic, American, and Korean alphabets. In the past decade, several research efforts have been made to automate sign language processing. Some researchers used traditional classifiers, while others used convolutional neural networks or recurrent neural networks (RNNs).

Concerning Arabic sign language (ArSL) recognition, Al-zohairi, Alghonaim, et al. [9], for example, presented an SVM model to classify ArSL images. The authors, using a smartphone camera, collected a dataset of 900 images for 30 alphabet characters and extracted features from images

using the histogram of oriented gradient (HOG) descriptor. The model achieved a low accuracy of 63.5%. In addition, Hasasneh and Taqatqa [14] proposed a model based on a restricted Boltzmann machine and tiny images for 39 Arabic alphabetic sign language groups.

Convolutional neural networks have also been applied for ArSL recognition. For example, Alani and Cosma [15] proposed two CNN models consisting of seven convolutional layers and four pooling layers for the ArSL 2018 dataset. The first one achieved an accuracy of 96.59%, while the second one used some sampling techniques to improve the accuracy to 97.29%. Also, Elsayed and Fathy [16] also implemented a CNN containing five convolutional layers and three pooling models for the Arabic alphabet and word recognition. They used a premade dataset consisting of 54049 samples to evaluate their model and produced a low accuracy of 88.87% for alphabet pattern recognition.

On the other hand, several techniques for American Sign Language (ASL) classification have been extensively studied. This usually consists of a two-stage feature extraction and a classifier. For instance, Aly, Aly, et al. [17] developed an SVM model to classify the ASL alphabet using a dataset collected by Microsoft Kinect from several users. The collected data is preprocessed for segmentation and feature extraction by the principal component analysis network (PCANet). This model achieves an accuracy of 88.7%. Shin, Matsuoka, et al. [12] proposed a model using SVM and light gradient boosting machine (GBM) to classify the ASL alphabets using the Massey alphabets dataset and the Kaggle alphabets dataset. The results were 99.39% for Massey and 87.60% for Kaggle.

The CNN technique is also used for ASL classification. For example, Fierro and Perez [8] built two CNN models for sharing parameters and achieved 96% accuracy. This model was fed by samples taken from the Kaggle dataset, which contains 29 subclasses. In addition, Abdulhussein and Raheem [13] also built a CNN model for classifying 24 ASL characters after preprocessing input images using image resizing, converting images to grayscale, and edge detection. Their model achieved an accuracy of 99.3%, and the training time was shorter compared to its peers. Furthermore, Wardana, Rachmawati, et al. [18] proposed a CNN model for the Kaggle ASL dataset, achieving an accuracy of 99.81%. The data is divided 70% for training, and the remainder of the dataset is equally divided between validation and testing. Also, Can, Kaya, et al. [19] suggested a CNN model for colored natural ASL pictures and compared their accuracy with five well-known transfer learning models, including VGG16, VGG19, ResNet50, and DenseNet121, to obtain 99.91% superior to his peers.

Finally, for Korean alphabets (KSL), Na, Yang, et al. [20] presented an SVM model for classifying 31 signs consisting of 14, 10, and 7 consonants, vowels, and double vowels, respectively. This dataset was collected from 15 participants who wore gloves while taking images. The tri-axial accelerometer signals were used to segment the sign gesture while the user was performing it. This model achieves a segmentation accuracy of 98.9%, which is superior to its peers, which used multiple sensors for segmentation. This model achieved a mean recognition accuracy of 92.2% for the Korean alphabet. Multiclass SVM was designed with six different kernels (e.g., linear, quadratic, and cubic) and optimized through

accuracy comparison. The quadratic kernel produced the best classification rate among the others. In contrast, Yeo and Shin [21] proposed a model for 12 classes of consonant and vowel letters. The model was designed by combining electromyography (EMG), accelerometers, and gyro sensors to build a multi-variate Gaussian model and maximum likelihood estimation through Bayesian theory for classification. As a result, accuracy rates of 99.13% and 99.97% were achieved for consonants and vowels, respectively.

Most researchers aimed to improve the performance of the CNN by studying the influence of preprocessing techniques such as filters [22] or other techniques (e.g., cropping, adding noise, and data normalization) [23]. Despite the preprocessing success in some cases, there is no way to generalize it for all datasets. Therefore, in the proposed model, preprocessing is not utilized to examine the power of the CNN, test the effect of its parameter in the original data, and test the model against dataset variation.

All the aforementioned traditional or modern techniques for sign language recognition models were trained and tested on a single dataset and ignored the diversity of various sign languages. In addition, some researchers used a small, collected dataset to train their models, so they could not be generalized. They also neglect prediction time, although it is a critical factor in real-time applications. Finally, even though the CNN technique was used, the CNN parameters were not studied sufficiently to achieve optimal accuracy and a reasonable prediction time. The main contribution of this paper can be summarized as follows: First, an efficient CNN sign language recognition is developed by studying the effect of CNN hyper-parameters to achieve higher accuracy compared with its peers using original data without any preprocessing. Second, the proposed model is trained on three sign language alphabets, namely, English, Arabic, and Korean, to evaluate its robustness against data variation. Third, the model is trained and tested to classify numerous sign languages by combining two different sign language datasets, namely Arabic and English. Finally, the prediction time is carefully considered during the design of the proposed model.

### III. METHODOLOGY OF CNN HYPER-PARAMETERS SELECTION

A CNN is one of the most popular techniques in deep learning (DL) that is successfully used in classification problems [24] and has high success rates in several problems such as hand gestures and object detection [25]. CNN consists of convolution, pooling, and fully connected layers. The convolution layers are the main layers in CNN, as they are responsible for creating a feature map using a set of filters whose number is a design parameter, or hyper-parameter [26]. Convolution is a linear multiplication operation between the input image and a filter whose size (the kernel size is another hyper-parameter) is smaller than the input image [27]. This results in an activation map (feature map) whose size is less than the input image and whose count is the same as the number of filters used in this operation. The size of the feature map is calculated according to Eq. 1.

$$output\ size = \frac{N - N_f}{stride} + 1 \quad (1)$$

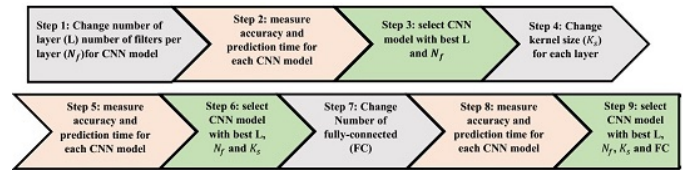


Fig. 1. Flow diagram of the methodology framework for the selection of CNN hyper-parameters.

where  $N$  is the width or height of the input image,  $N_f$  is the filter's width or height, and the stride is the pixel size between each convolution [28]. The size of the feature map in Eq. 1 must be an integer number; otherwise, padding is employed. Padding is the process of increasing the image's size without changing its content. The input image, after being convolved repeatedly with filters, shrinks in volume spatially. Shrinking too fast is not good; it does not work well. Pooling combines the nearby units to reduce the input size for the next layer. It includes maximum pooling and average pooling. Using the pooling layer directly after the convolution layer is not necessary, but its type and location are also a hyper-parameter whose settings are set empirically using expertise. The result of repeated convolution and pooling is that the list of features (a vector of features) is the input for the last layer of CNN (the number of convolutional layers is a hyper-parameter). The fully connected layer classifies the input to the best class [29].

The multiple hyper-parameters that CNN uses to make the process of picking an architecture much more challenging [30]. In this section, a detailed study is conducted for the optimal selection of CNN model hyper-parameters, namely kernel size ( $K_s$ ), the number of filters ( $N_f$ ), the number of convolution layers ( $L$ ), and the number of fully connected hidden layers, to achieve optimal accuracy in a reasonably short time for ArSL. The steps of the methodology are summarized in Fig. 1. Initially, this method looked at the effect of varying the layer and filter numbers on precision. Secondly, a model is selected based on its speed and accuracy of prediction. Finally, the influence of kernel size and the number of hidden layers is considered to select the optimal CNN architecture. The methodology is discussed in detail in the following subsections.

#### A. Selecting the Number of Layers and Filters Per Layer

The number of filters  $N_f$  in each convolution layer affects the test time because the more filters used, the more computational time is required. This occurs because the output of a convolution layer, i.e., the number of activations maps, equals the number of filters used in that layer [31]. The number of filters to be used in each layer is chosen according to Eq. 2.

$$N_f = 2^k; \quad k = 2, 3, 4, 5, \dots \quad (2)$$

to achieve the best accuracy within a reasonable prediction time. Also, the number of convolution layers has a vital role in the classification time in the ArSL problem. Because the image size was 64 by 64, the maximum number of convolutions with stride 2 would be 6. So, the number of layers is varied between 1 to 6, and tests are repeated 50 times to determine the best number of layers that gives the best average test time.



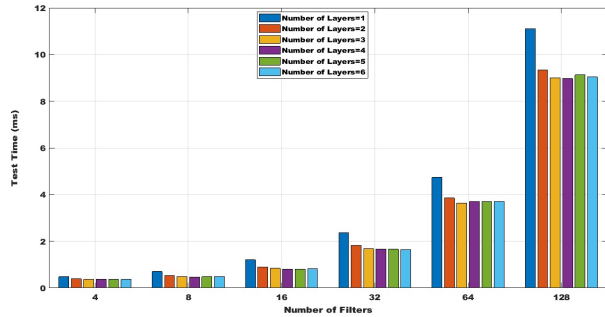


Fig. 2. Test time as a function of the number of layers and number of filters per layer.

TABLE I. ACCURACY AND TEST TIME FOR ARSL USING DIFFERENT NUMBERS OF LAYERS AND FILTERS PER LAYER

L	$N_f=4$		$N_f=16$		$N_f=128$	
	Acc(%)	$T_S$ (ms)	Acc (%)	$T_S$ (ms)	Acc (%)	$T_S$ (ms)
L = 2	95.4770	0.3793	96.8450	0.8819	96.586	8.9003
L = 4	95.2921	0.3606	97.6707	0.7986	98.52	8.8196
L = 6	56.6182	0.3598	96.1794	0.7969	98.5827	8.9013

Acc is the accuracy and  $T_S$  is the Test time in milliseconds //

In this study, several CNN models are built with different combinations of layers and filters per layer. The test time is estimated without training, as shown in Fig. 2. It can be realized that the test time increases with the number of filters per layer but decreases as the number of layers increases. This is because the number of extracted features from CNN (the last layer of CNN) decreases.

On the other hand, to take accuracy into account and decide the optimal number of layers and filters per layer, the CNN models with all combinations of layer numbers 2, 4, and 6 and filter numbers 4, 16, and 128 were trained. Table I shows the test time and accuracy in classifying the ArSL dataset. As can be seen from Table I, although a CNN with 4 filters per layer has the least amount of time, its accuracy is inferior compared to that obtained using 16 or 128 filters. Therefore, it is clear that the number of filters per layer should be at least 16. Considering the number of layers, it is noted that accuracy obtained using more than 4 layers either drops ( $N_f = 4$  or 16) or is not significantly improved ( $N_f = 128$ ). Based on these observations, the number of layers in the proposed model is set to 4. For this number of layers, the accuracy improved with the increase in the number of filters. Unfortunately, this comes at the expense of a significantly longer prediction time. Therefore, in the proposed model, instead of using 128 filters in all layers, the number of filters in the four layers is set respectively to 32, 64, 128, and 128. This reduces the total prediction time while maintaining acceptable accuracy at the last two layers.

### B. Selecting the Kernel Size

Kernel size, which is usually taken as an odd number less than 10, also affects the features produced by the convolution [32]. Therefore, kernel sizes of 3, 5, 7, and 9 are trained for the ArSL classification problem. During this experiment, 10

TABLE II. TRAINING, TEST ACCURACY, TRAINING, AND TESTING TIME VS. CNN KERNEL SIZE

$K_s$	$T_r$ Acc (%)	$T_s$ Acc (%)	$T_r$ time (h)	$T_S$ time (ms)	Std ( $T_s$ )
3x3	100	98.55	3.39	2.6621	$4.4920 \times 10^{-5}$
5x5	100	98.71	5.02	3.5870	$1.1804 \times 10^{-4}$
7x7	100	98.87	6.62	8.0030	$1.0546 \times 10^{-4}$
9x9	100	98.77	9.27	11.8441	$1.5197 \times 10^{-4}$

$T_r$  is training,  $T_s$  is test and std is the Standard deviation.

TABLE III. TEST ACCURACY, TRAINING AND TESTING TIME VS. CNN FULLY CONNECTED HIDDEN LAYER

Hidden No.	$T_r$ Acc(%)	$T_s$ Acc(%)	$T_r$ time(h)	$T_s$ time(ms)	std( $T_s$ )
0	100	98.3362	4.155	2.8157	$3.0857 \times 10^{-5}$
1	100	98.5457	4.385	2.8760	$3.3840 \times 10^{-5}$
2	100	98.7183	4.360	2.8893	$3.4234 \times 10^{-5}$
3	100	98.5580	4.455	2.9107	$1.7433 \times 10^{-5}$

$T_r$  is training,  $T_s$  is test and std is the Standard deviation.

runs are performed, and the average results are reported in Table II. As it can be seen, kernel size has a slight effect on the accuracy of training and testing. In contrast, the training time and test time increase nonlinearly with kernel size. Based on these observations, the optimal kernel size would be 3x3, as it achieves the minimum training and testing times while maintaining the same accuracy.

### C. Selecting the Number of Fully Connected Hidden Layers

CNN always ends with a fully connected network that contains hidden layers whose number is considered a hyper-parameter [33]. The network with 0, 1, 2, and 3 hidden layers is trained in ArSL to investigate the best number of hidden layers. Each model is trained and tested ten times, and the average results are reported in Table III. As can be noted, fully connected two hidden layers achieve the best test accuracy with slightly higher test and training times. By adding the last output layer to the two hidden layers, three fully connected layers are considered in the proposed CNN model.

### D. Study the Time Complexity of CNN's Prediction Time

In this section, the time complexity of CNN's prediction time is derived. The number of operations in a single convolution layer depends on the size of the image ( $N \times N$ ), the kernel size and the number of filters per layer. The number of convolutions is performed for each pixel in the image, so the time complexity is proportional to the total number of pixels in the image ( $N^2$ ). Each pixel is multiplied by a window of size  $K_s \times K_s$ . This is done for each filter; so, the time complexity for a single convolution layer is given by Eq. 3.

$$T_{conv} = O(N^2 K_s^2 N_f) \quad (3)$$

The feature map output from a single convolution layer is equal to the image multiplied by the number of filters in this layer. To reduce the dimension of the feature map, a Max-pooling layer is applied with a 2x2 window in the feature map of size  $N \times N \times N_f$ , whose time complexity can be expressed as Eq. 4.

$$T_{mp} = O(4N^2 N_f) \quad (4)$$

For L layers, the time complexity in Eq. 4 can therefore be written as Eq. 5.

$$T_L = O((N^2 K_s^2 N_f + 4N^2 N_f)L) \quad (5)$$

where time complexity is linear in L. For fully connected layers, the feature maps are generated from the final convolution layer and the size of the input layer. The max-pooling layer decreases the size of the image by one-fourth (for a stride of 2). So, the final feature map size is  $N^2/4^L$ , and the time complexity of the first fully connected layer would be Eq. 6.

$$T_{FC} = O\left(\frac{N^2 N_f}{4^L}\right) \quad (6)$$

So, the total time complexity would be calculated as shown in Eq. 7.

$$O((N^2 K_s^2 N_f + 4N^2 N_f)L + \frac{N^2 N_f}{4^L} + TimeofOtherFC) \quad (7)$$

The time for other fully connected layers is dropped in big-O notation from Eq. 7 as they are constant terms. Therefore, the time of fully contented dropped and all constants are dropped as shown in Eq. 8.

$$T = O(K_s^2 N^2 N_f L + \frac{N^2 N_f}{4^L}) \quad (8)$$

From Eq. 8, it can be concluded that the time complexity of the CNN model is linearly proportional to Nf, the square of Ks, and nonlinear with L. Furthermore, to investigate the validity of Eq. 8, three examinations are made by building several CNN models, changing one parameter at a time, and averaging the test time over 50 runs. The procedure can be stated as follows:

- Varying the  $K_s$  from 1 to 19, while fixing the layer number at 4 and the number of filters at 16. Fig. 3 shows the measured test time as a function of the  $K_s$ . As it can be seen, the relationship can be fitted using a quadratic polynomial, which means that the time complexity of the CNN model is proportional to the square of  $K_s$ .
- Investigating the effect of the number of layers L on test time (at  $K_s = 3, N_f = 16$ ). Fig. 4 shows that the time decreases as the number of layers increases up to 4 layers. Above 4 layers, the time starts to increase slightly. This experiment shows that the test time is inversely proportional to the number of layers, and a good fit can be obtained using a fifth-degree polynomial.
- Examining the effect of the number of filters Nf on test time (at  $K_s = 3, L=4$ ). Fig. 5 shows that the test time is linearly proportional to the number of filters and can be fitted using a linear equation.

From Eq. 8 and the previous experiments of CNN, it is noted that the time complexity of the CNN model increases as the kernel size and number of filters increase. Moreover, the  $K_s$  significantly affects the test time because it is in proportion to the square of the  $K_s$ . On the other hand, the test time decreases as the number of layers increases.

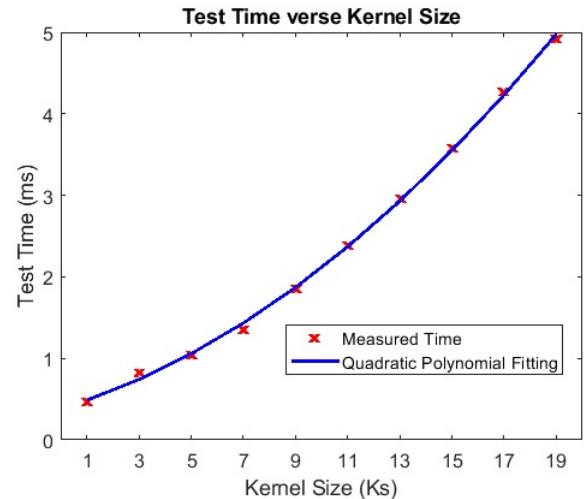


Fig. 3. Test time verse CNN  $K_s$ .

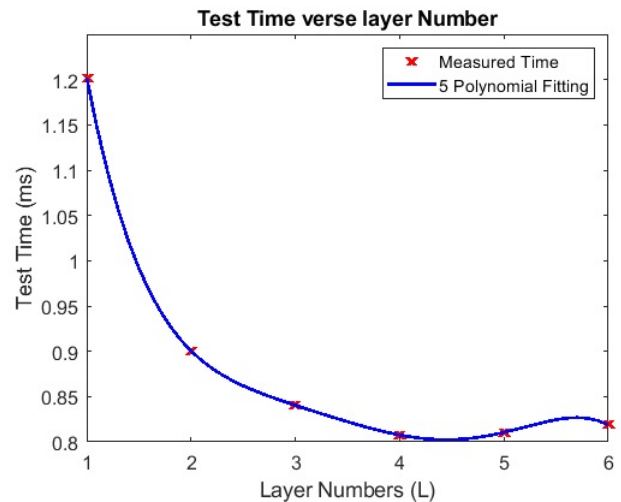


Fig. 4. Test time verse CNN layer numbers (L).

#### IV. THE PROPOSED SL-CNN MODEL

In this section, the architecture of the proposed model is designed to achieve high performance and less prediction time. Also, to make the model operate directly on raw data to generalize it across several datasets and data robustness. The architecture of the proposed SL-CNN model is presented next and summarized as shown in Fig. 6. Based on the analysis of the experimental results in Section III, the proposed SL-CNN model is designed using four layers with a kernel size of 3 x 3. In addition, filter numbers ranging from 32 to 128 were also tested to achieve optimal accuracy and prediction times. The proposed SL-CNN model consists of four convolutional (Conv) layers, four max pooling (MP) layers, three fully connected layers, and two dropout layers, as shown in Fig. 7. The pooling layer has a Relu activation function. Additionally, 7 batch normalization (BN) layers are used to achieve a stable distribution of the activation values through training and normalizing the input layers [34]. Also, each convolution layer is followed by a Relu function layer.





Fig. 8. Sample signs from the ArSL 2018 Arabic sign language dataset.

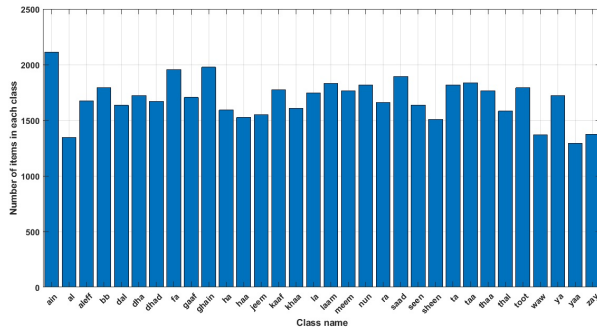


Fig. 9. Number of images in each class in ArSL dataset.

a) *Kaggle dataset (ASL 2018)*: The 2018 update of the American Sign Language (ASL) alphabet dataset, available on Kaggle’s data science repository, contains 29 classes for alphabet characters. Each class contains 3000 images captured with the intent to make a dataset for each character [37]. The dataset is divided into 70% for training, and the remainder is divided equally between validation and testing. Table IV shows the number of images for each partition. The data is rearranged randomly for each experiment.

b) *Massey dataset (ASL 2012)*: The last version of the Massey dataset, introduced in January 2012, contains 1815 images of 26 ASL alphabet gestures [38]. Each class contains 70 images, so the data is balanced. The dataset is divided into 85% for training, and the remainder is divided equally between validation and testing, as shown in Table IV. The data is rearranged randomly for each experiment.

3) *Arabic-English dataset*: The Arabic-English dataset combines the ArSL 2018 and ASL 2018 datasets, and the model deals with them as one dataset. The dataset contains 61 classes; each class has a different number of images, dimensions, and colors. Fig. 10 shows the number of images in each class. The combined dataset is divided into 70% for training, and the remainder is divided equally between validation and testing, as shown in Table IV.

4) *KSL consonant letters*: Korean consonant letters are also available on Kaggle and were last updated in June 2021. It contains 14 classes for consonant Korean alphabet characters [39]. The data is supported by 21962 images for training and validation and 3790 for the test. The images in the

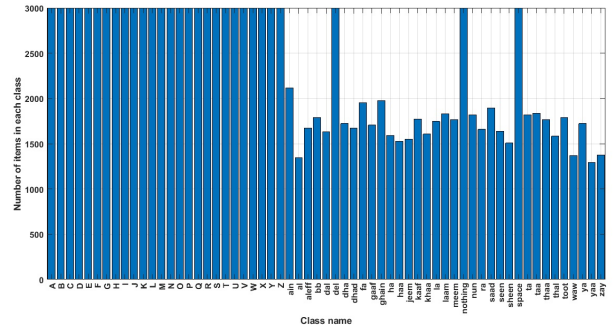


Fig. 10. Number of images in each class in the Arabic-English dataset.

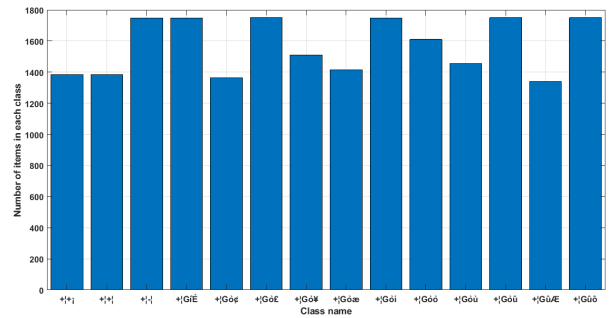


Fig. 11. Number of images in each class in the KSL dataset.

TABLE IV. DATA DIVISION BETWEEN TRAINING AND TESTING

Dataset	Total images	training images	val images	test images
ArSL	54049	37835	8110	8114
ASL2018	87000	60900	13050	13050
ASL2012	1815	1555	130	130
Arabic-English	141049	98735	21150	21164
KSL	25752	18671	3291	3790

Korean dataset are imbalanced distributed, as shown in Fig. 11. Training data is divided into 85% for training and 15% for validation, as shown in Table IV. Data is randomly fed to the network at each iteration.

### B. Experimental Setup

The experiment was conducted using MATLAB® 2020 Deep Learning Toolbox running on a 2.60 GHz Intel i5 CPU with 8 GB RAM, Intel 4 K graphics, and AMD Radeon 7500M/7600M series.

To evaluate the robustness of the model against randomization and avoid bias towards certain parameters, the experiment was conducted ten times; in each trial, the dataset was randomly divided into training and testing, and the results were averaged [15]. Also, the dropout layer is added to avoid the overfitting problems [40]. The effectiveness of the proposed model is evaluated on five datasets: (1) the ArSL 2018 dataset; (2) the ASL 2018 dataset; (3) the ASL 2012 dataset; (4) the ArSL-ASL combination; and (5) the KSL 2021 dataset. The accuracy defined as Eq. 9 is used to evaluate the classification performance, where FC and TC denote the total number of

TABLE V. CLASSIFICATION ACCURACY, TRAINING AND TEST TIME

Dataset	$T_r$ Acc(%)	$T_s$ Acc(%)	$T_r$ time(h)	$T_s$ time(ms)	std for $T_s$
ArSL 2018	100	98.7799	3.93	2.6698	0.1436
ASL 2018	100	100.00	6.46	3.1746	0.1433
ASL 2012	100	100.00	0.2558	1.4912	0.4876
ArSL-ASL	100	99.4897	10.5	3.0440	0.4066
KSL 2021	100	100.00	2.60	2.7319	0.0737

where  $T_r$  means training and  $T_s$  means test

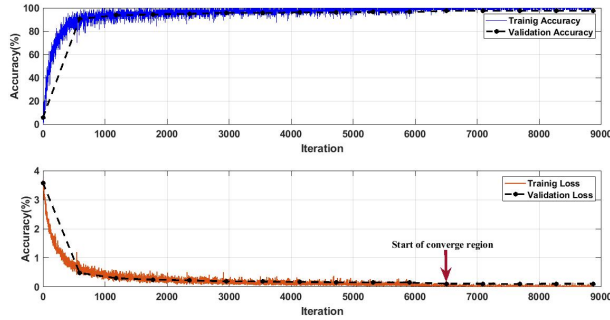


Fig. 12. Training and validation accuracy for ArSL.

false and correct instances of the test. Also, the sensitivity for each class is expressed as Eq. 10 and is used to express the evaluation of the performance for each class individually, where  $TC_c$  and  $TF_c$  are the numbers of correct and false instances in each class.

$$Accuracy = \frac{TC}{FC + TC} * 100 \quad (9)$$

$$TC_c = \frac{TC_c}{FC_c + TC_c} * 100 \quad (10)$$

### C. Performance Evaluation of the Proposed SL-CNN Model

The proposed model was trained for 15 or 20 epochs, and the proposed algorithm achieves high accuracy at epoch number 15. Table V summarizes the training accuracy, test accuracy, training time, test time, and standard deviation for test time with 15 epochs for the proposed SL-CNN model on ArSL 2018, ASL 2018, ASL 2012, the combination between them (ArSL-ASL) datasets, and 20 epochs for KSL 2021. Fig. 12 to 16 show the training and test accuracies and the start of the loss function to converge for ArSL 2018, ASL 2018, ASL 2012, ArSL-ASL, and KSL, respectively. It is noted from the table that the training accuracy for all datasets is 100%, and the test accuracy is 100% for ASL 2018, ASL 2012, and KSL, as well as 98.8% and 99.5% for ArSL and the combined dataset. In addition, the training and test times are different from one dataset to another. Figs. 17, 18, 19, and 20 show the confusion matrices for ArSL, ASL 2018, ASL 2012, and KSL. The diagonal of the confusion matrix refers to the number of the correct element obtained by the model, while off-diagonal predictions are false.

### D. Comparison with Other Models

In this section, we make a comparison between the SL-CNN model and other models published in other articles. The

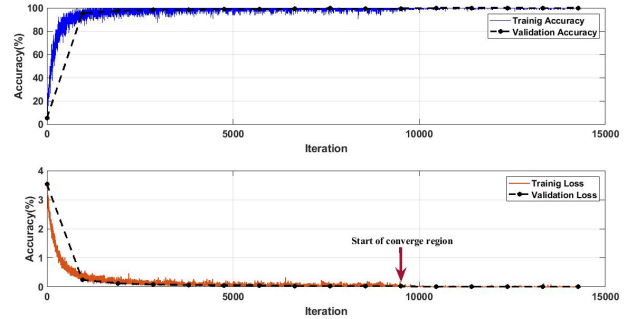


Fig. 13. Training and validation accuracy for ASL 2018.

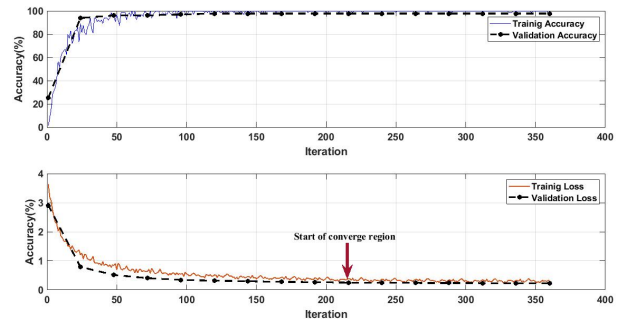


Fig. 14. Training and validation accuracy for ASL 2012.

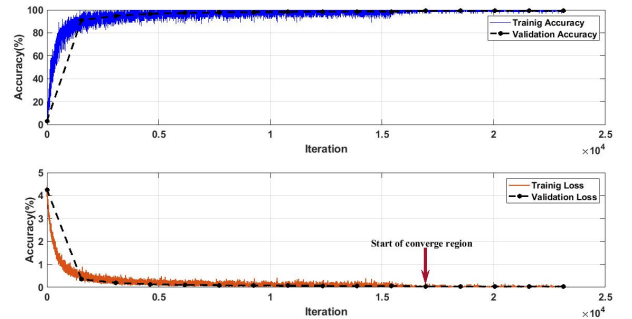


Fig. 15. Training and validation accuracy for ArSL-ASL.

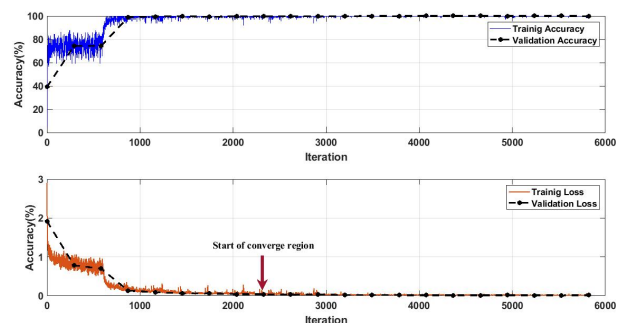


Fig. 16. Training and validation accuracy for KSL.

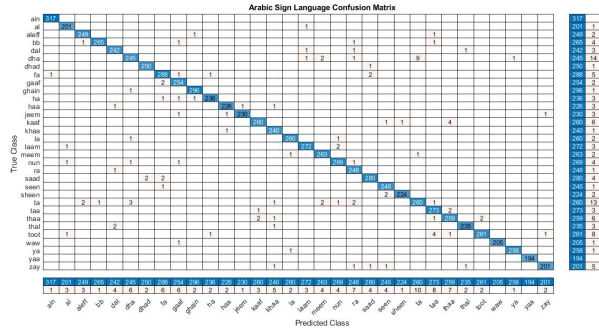


Fig. 17. Confusion matrix for ArSL.

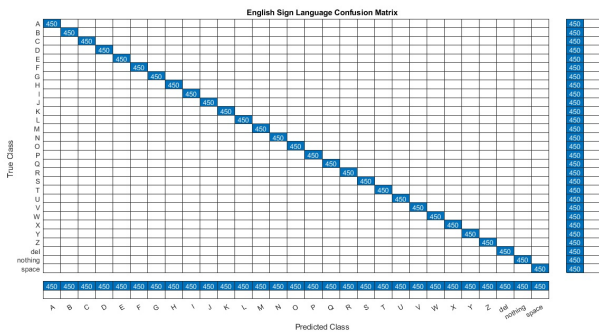


Fig. 18. Confusion matrix for ASL 2018.

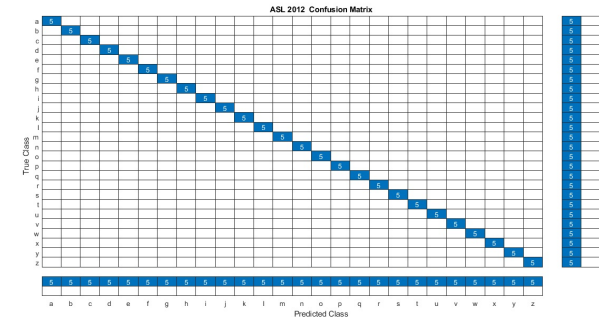


Fig. 19. Confusion matrix for ASL 2012.

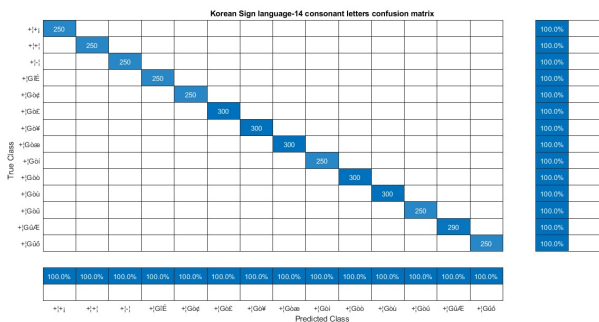


Fig. 20. Confusion matrix for KSL.

TABLE VI. COMPARISON OF THE RESULTS BETWEEN THE PROPOSED MODEL AND OTHER PREVIOUS ARTICLES FOR ARSL

Author	Method	Dataset	Samples	Acc(%)
Latif, et al. [41]	CNN	ArSL 2018	54049	97.6
Elsayed, et al. [16]	CNN	ArSL 2018	54049	88.89
Alani, et al. [15]	CNN	ArSL 2018	54049	97.29
proposed SL-CNN	CNN	ArSL 2018	54049	98.47
M. Zakariah, et al. [42]	EfficientNetB4	ArSL 2018	54049	95.0
Rehab, et al. [43]	VGGNet	ArSL 2018	54049	97.0

comparison is divided into three parts ArSL, ASL, and KSL. It is noted that the proposed SL-CNN model outperforms its peers because each parameter of the CNN model structure has been carefully selected, as mentioned in Section III, and doesn't depend on preprocessing techniques to improve the accuracy of the application like others; but it works on raw data. For the ArSL 2018 dataset, the proposed SL-CNN model achieves 98.47% accuracy, outperforming other CNN models proposed in the literature, e.g. [16], [41] and [15] as shown in Table VI. It should be noted that the proposed model uses raw data without any preprocessing. It should be highlighted that the models by [42] and [43] employed modern classification approaches (Transfer learning) but obtained poor accuracy when compared to the proposed SL-CNN model.

For the ASL, the proposed model is compared to the models proposed in Kaggle datasets [8], [12], [18]] and [[12], [44], [45]] on the Massey dataset, which is shown in Table VII. The result of other state-of-the-art in literature [17] and [13] on collected but not publicly available datasets are also reported in Table VII. The CNN models proposed in [[46], [19]] achieved higher accuracy on selected images from the MNIST dataset. However, the proposed SL-CNN model is applied to all images in the recent Kaggle dataset, and the two versions of Massey beat all the previously suggested CNN models. Interestingly, the proposed model not only outperforms published models on ASL 2018 and ASL 2012 but also achieves 100% accuracy, in accordance with the recent results published on Kaggle [37]. It should be noted that the model by [47] achieved also 100% accuracy, however, on the Massey dataset 2011. Also, the model proposed in [13] got 99.3 % accuracy for ASL classification, but they used a collected dataset consisting of 240 images only.

Finally, for the Korean sign language, state-of-the-art methods in the literature are performed on collected but not publicly available datasets, as reported in VIII. The proposed model achieved 100% test accuracy for consonant characters, compared to the models by [20] who achieved 92.2% for vowel and consonant characters, and [21] who achieved 99.31% and 99.97% for vowel and consonant characters, respectively.

## VI. CONCLUSION AND FUTURE WORK

This paper proposes an approach to designing and analyzing an efficient CNN model for sign language recognition named SL-CNN based on a detailed study of CNN hyper-parameters, i.e., kernel size, number of layers, and filtering number in each layer. The performance of the proposed model was investigated for four sign language datasets (ArSL 2018, ASL 2018, ASL 2012, and KSL 2021), and one dataset was

TABLE VII. COMPARISON OF THE RESULTS BETWEEN THE PROPOSED MODEL AND OTHER PREVIOUS ARTICLES FOR ASL

Table with 5 columns: Author, Method, Acc (%), Dataset, Samples. Rows include Aly et al. [17], Raheem et al. [13], Taskiran et al. [47], A. Mannan et al. [46], C. Can et al. [19], Shin et al. [12], Rastgoo et al. [44], Rahman et al. [45], Proposed model, Shin et al. [12], Wardana et al. [18], Fierro et al. [8], and Proposed model.

TABLE VIII. COMPARISON OF THE RESULTS BETWEEN THE PROPOSED MODEL AND OTHER PREVIOUS ARTICLES FOR KSL

Table with 5 columns: Author, Method, Dataset, Classes, Acc (%). Rows include [20], [21], and Proposed.

obtained as a merge of the first two datasets (ArSL and ASL 2018). Table VIII shows the comparison of the results between the proposed model and other previous articles for KSL. The results show that the proposed SL-CNN model outperforms the other models in terms of classification accuracy for all datasets. In summary, the proposed model achieved accuracy for 98.8%, 100%, 99.5%, 100% and 100% for ArSL2018, ASL 2018, combining both, ASL 2012 and KSL 2021, respectively without resorting to any data preprocessing. As suggestions for future work, it is interesting to investigate the performance of the proposed model on sign language datasets consisting of words and sentences. Another possibility is to implement the proposed model in, e.g., an educational system for people who suffer from hearing problems.

ACKNOWLEDGMENT

Special thanks to Dr. Ahmed Mahmoud Alenany and Dr. Mohamed L. Elsayed for the English review of this manuscript.

REFERENCES

[1] E. McPhillips, "World wide hearing loss: Stats from around the world," Audicus. Retrieved September, vol. 10, p. 2022, 2022.
[2] M. Mustafa, "A study on arabic sign language recognition for differently abled using advanced machine learning classifiers," Journal of Ambient Intelligence and Humanized Computing, vol. 12, pp. 4101-4115, 2021.

[3] C. K. Lee, K. K. Ng, C.-H. Chen, H. C. Lau, S. Chung, and T. Tsoi, "American sign language recognition and training method with recurrent neural network," Expert Systems with Applications, vol. 167, p. 114403, 2021.
[4] S.-K. Ko, J. G. Son, and H. Jung, "Sign language recognition with recurrent neural network using human keypoint detection," in Proceedings of the 2018 conference on research in adaptive and convergent systems, 2018, pp. 326-328.
[5] R. Rastgoo, K. Kiani, and S. Escalera, "Sign language recognition: A deep survey," Expert Systems with Applications, vol. 164, p. 113794, 2021.
[6] E.-S. M. El-Alfy and H. Luqman, "A comprehensive survey and taxonomy of sign language research," Engineering Applications of Artificial Intelligence, vol. 114, p. 105198, 2022.
[7] A. Al-Naji, K. Gibson, S.-H. Lee, and J. Chahl, "Real time apnoea monitoring of children using the microsoft kinect sensor: a pilot study," Sensors, vol. 17, no. 2, p. 286, 2017.
[8] K. R. Perez-Daniel and A. N. Fierro Radilla, "Siamese convolutional neural network for asl alphabet recognition," OPENAIRE, 2020.
[9] R. Alzohairi, R. Alghonaim, W. Alshehri, and S. Aloqeely, "Image based arabic sign language recognition system," International Journal of Advanced Computer Science and Applications, vol. 9, no. 3, 2018.
[10] M. Abdo, A. Hamdy, S. Salem, and E. M. Saad, "Arabic alphabet and numbers sign language recognition," International Journal of Advanced Computer Science and Applications, vol. 6, no. 11, pp. 209-214, 2015.
[11] C. Dong, M. C. Leu, and Z. Yin, "American sign language alphabet recognition using microsoft kinect," in Proceedings of the IEEE conference on computer vision and pattern recognition workshops, 2015, pp. 44-52.
[12] J. Shin, A. Matsuoka, M. A. M. Hasan, and A. Y. Srizon, "American sign language alphabet recognition by extracting feature from hand pose estimation," Sensors, vol. 21, no. 17, p. 5856, 2021.
[13] A. A. Abdulhusein and F. A. Raheem, "Hand gesture recognition of static letters american sign language (asl) using deep learning," Engineering and Technology Journal, vol. 38, no. 6, pp. 926-937, 2020.
[14] N. Hasasneh, A. Hasasneh, and S. Taqatqa, "Towards arabic alphabet and numbers sign language recognition," 2017.
[15] A. A. Alani and G. Cosma, "Arsl-cnn: a convolutional neural network for arabic sign language gesture recognition," Indonesian journal of electrical engineering and computer science, vol. 22, 2021.
[16] E. K. Elsayed and D. R. Fathy, "Sign language semantic translation system using ontology and deep learning," International Journal of Advanced Computer Science and Applications, vol. 11, no. 1, 2020.
[17] W. Aly, S. Aly, and S. Almotairi, "User-independent american sign language alphabet recognition based on depth image and pcanet features," IEEE Access, vol. 7, pp. 123 138-123 150, 2019.
[18] B. K. Wardana, E. Rachmawati, and T. A. B. Wirayuda, "Pengenalan gestur tangan statis menggunakan cnn dengan arsitektur efficient-net b4," eProceedings of Engineering, vol. 8, no. 2, 2021.
[19] C. Can, Y. Kaya, and F. Kılıç, "A deep convolutional neural network model for hand gesture recognition in 2d near-infrared images," Biomedical Physics & Engineering Express, vol. 7, no. 5, p. 055005, 2021.
[20] Y. Na, H. Yang, and J. Woo, "Classification of the korean sign language alphabet using an accelerometer with a support vector machine," Journal of Sensors, vol. 2021, pp. 1-10, 2021.
[21] I. Yeo and H.-C. Shin, "Novel korean finger language recognition using emg and motion sensors," in 2018 International Conference on Information Networking (ICOIN). IEEE, 2018, pp. 837-839.
[22] A. H. Barshooi and A. Amirkhani, "A novel data augmentation based on gabor filter and convolutional deep learning for improving the classification of covid-19 chest x-ray images," Biomedical Signal Processing and Control, vol. 72, p. 103326, 2022.
[23] D. A. Pitaloka, A. Wulandari, T. Basaruddin, and D. Y. Liliana, "Enhancing cnn with preprocessing stage in automatic emotion recognition," Procedia computer science, vol. 116, pp. 523-529, 2017.
[24] H. Moujahid, B. Cherradi, O. El Gannour, W. Nagmeldin, A. Abdelmaboud, M. Al-Sarem, L. Bahatti, F. Saeed, and M. Hadwan, "A novel explainable cnn model for screening covid-19 on x-ray images," Comput. Syst. Sci. Eng., vol. 46, no. 2, pp. 1789-1809, 2023.

- [25] A. Amirkhani, A. H. Barshooi, and A. Ebrahimi, "Enhancing the robustness of visual object tracking via style transfer." *Computers, Materials & Continua*, vol. 70, no. 1, 2022.
- [26] R. Garg, S. Maheshwari, and A. Shukla, "Decision support system for detection and classification of skin cancer using cnn," in *Innovations in Computational Intelligence and Computer Vision: Proceedings of ICICV 2020*. Springer, 2021, pp. 578–586.
- [27] S. Wen, J. Chen, Y. Wu, Z. Yan, Y. Cao, Y. Yang, and T. Huang, "Ckfo: Convolution kernel first operated algorithm with applications in memristor-based convolutional neural network," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 40, no. 8, pp. 1640–1647, 2020.
- [28] F. F. Li, J. Justin, and S. Yeung, "Course notes on cs231n: Convolutional neural networks for visual recognition," 2018.
- [29] H.-C. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Noguees, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.
- [30] M. Coletti, D. Lunga, J. K. Bassett, and A. Rose, "Evolving larger convolutional layer kernel sizes for a settlement detection deep-learner on summit," in *2019 IEEE/ACM Third Workshop on Deep Learning on Supercomputers (DLS)*. IEEE, 2019, pp. 36–44.
- [31] W. S. Ahmed *et al.*, "The impact of filter size and number of filters on classification accuracy in cnn," in *2020 International conference on computer science and software engineering (CSASE)*. IEEE, 2020, pp. 88–93.
- [32] J. Brownlee, "A gentle introduction to padding and stride for convolutional neural networks," *Machine Learning Mastery*, 2019.
- [33] S. Shin, Y. Lee, M. Kim, J. Park, S. Lee, and K. Min, "Deep neural network model with bayesian hyperparameter optimization for prediction of nox at transient conditions in a diesel engine," *Engineering Applications of Artificial Intelligence*, vol. 94, p. 103761, 2020.
- [34] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*. pmlr, 2015, pp. 448–456.
- [35] K. Alkharabsheh, S. Alawadi, V. R. KEBande, Y. Crespo, M. Fernández-Delgado, and J. A. Taboada, "A comparison of machine learning algorithms on design smell detection using balanced and imbalanced dataset: A study of god class," *Information and Software Technology*, vol. 143, p. 106736, 2022.
- [36] L. Ghazanfar, A. Jaafar, M. Nazeeruddin, A. Roaa, and A. Rawan, "Arabic alphabets sign language dataset (arasl)," *Mendeley Data*, vol. 1, p. 2018, 2018.
- [37] AKASH. Asl alphabet. [Online]. Available: <https://www.kaggle.com/grassknotted/asl-alphabet>
- [38] A. Barczak, N. Reyes, M. Abastillas, A. Piccio, and T. Susnjak, "A new 2d static hand gesture colour image dataset for asl gestures," 2011.
- [39] BIO. Korean sign language-14 consonant letters. [Online]. Available: <https://www.kaggle.com/tr9904ewhaincom/korean-sign-language14-consonant-letters/activity>
- [40] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.
- [41] G. Latif, N. Mohammad, R. AlKhalaf, R. AlKhalaf, J. Alghazo, and M. Khan, "An automatic arabic sign language recognition system based on deep cnn: an assistive system for the deaf and hard of hearing," *International Journal of Computing and Digital Systems*, vol. 9, no. 4, pp. 715–724, 2020.
- [42] M. Zakariah, Y. A. Alotaibi, D. Koundal, Y. Guo, M. Mamun Elahi *et al.*, "Sign language recognition for arabic alphabets using transfer learning technique," *Computational Intelligence and Neuroscience*, vol. 2022, 2022.
- [43] R. M. Duwairi and Z. A. Halloush, "Automatic recognition of arabic alphabets sign language using deep learning," *International Journal of Electrical & Computer Engineering (2088-8708)*, vol. 12, no. 3, 2022.
- [44] R. Rastgoo, K. Kiani, and S. Escalera, "Multi-modal deep hand sign language recognition in still images using restricted boltzmann machine," *Entropy*, vol. 20, no. 11, p. 809, 2018.
- [45] M. M. Rahman, M. S. Islam, M. H. Rahman, R. Sassi, M. W. Rivolta, and M. Aktaruzzaman, "A new benchmark on american sign language recognition using convolutional neural network," in *2019 International Conference on Sustainable Technologies for Industry 4.0 (STI)*. IEEE, 2019, pp. 1–6.
- [46] A. Mannan, A. Abbasi, A. R. Javed, A. Ahsan, T. R. Gadekallu, Q. Xin *et al.*, "Hypertuned deep convolutional neural network for sign language recognition," *Computational intelligence and neuroscience*, vol. 2022, 2022.
- [47] M. Taskiran, M. Killioglu, and N. Kahraman, "A real-time system for recognition of american sign language by using deep learning," in *2018 41st international conference on telecommunications and signal processing (TSP)*. IEEE, 2018, pp. 1–5.



# Enhancing Outdoor Mobility and Environment Perception for Visually Impaired Individuals Through a Customized CNN-based System

Athulya N K<sup>1</sup>, Sivakumar Ramachandran<sup>2</sup>, Neetha George<sup>3</sup>, Ambily N<sup>4</sup>, Linu Shine<sup>5\*</sup>

Department of Electronics and Communication, College of Engineering Trivandrum, India<sup>1</sup>

Department of Electronics and Communication, Government Engineering College Wayanad, India<sup>2</sup>

Department of Electronics and Communication, Rajeev Gandhi Institute of Technology Kottayam, India<sup>3,5</sup>

Department of Electronics and Communication, Government Engineering College Idukki, India<sup>4</sup>

**Abstract**—Visual impairment indicates any kind of vision loss including blindness. Individuals with visual impairments face significant challenges when trying to perceive their surroundings from a global perspective and navigating unfamiliar environments. Existing assistive technologies predominantly focus on obstacle avoidance, neglecting to provide comprehensive information about the overall environment. To address this gap, the proposed system employs a customized Convolutional Neural Network (CNN) model tailored to accurately predict the type of outdoor ground terrain the user is traversing. This information is then conveyed to the user audibly. It can also detect the presence of puddles on the road and let the user know whether the outside floor is wet (slippery). The proposed deep-learning architecture is trained on images collected from sources including the Stagnant Water dataset, the GTOS-Mobile dataset and a custom dataset. The trained model is then integrated into an Android app, providing visually impaired (VI) people with effective surrounding perception capabilities, leading to better travel and, ultimately, better living.

**Keywords**—Visually impaired; terrain identification; puddle detection; deep learning

## I. INTRODUCTION

Visual impairment encompasses a range of conditions that can result in partial or complete loss of vision, including color blindness, affecting over 250 million people worldwide. Such individuals encounter difficulties while comprehending and interacting with their environment due to their limited visual contact with their surroundings. Physically moving around can be particularly challenging for them as they struggle to identify their location and navigate to different places.

For navigation, individuals with visual impairments often utilize white canes or guide dogs, although the latter option can be costly. However, despite the assistance provided by white canes, the task of maneuvering through unfamiliar environments continues to pose a significant challenge. Conventional navigation techniques primarily focus on obstacle avoidance [1], which curtails the capacity of visually impaired (VI) individuals to engage with novel surroundings fully. Consequently, many still face obstacles in gaining a holistic perception of their environment.

Recent advancements in technology have paved the way for the development of intelligent or augmented white canes

that utilize image processing and deep learning techniques to aid individuals with low vision in obstacle avoidance, object detection [2], and indoor/outdoor navigation [3]. However, these systems have a limitation in that they do not provide a comprehensive perception of the surrounding environment. For instance, while these canes assist in identifying objects and navigating outdoors, they cannot provide information on the terrain, which is crucial for individuals with visual impairments to determine the floor they are walking on. By incorporating terrain information into these systems, VI individuals can benefit from a more comprehensive and seamless travel experience, ultimately leading to an improved quality of life. This is particularly important for elderly individuals with low vision.

The proposed method makes significant contributions to assistive technology for VI individuals, with a strong focus on human needs and safety. The main contributions are:

- Proposed a novel custom CNN for Comprehensive Terrain Identification: An innovative lightweight CNN has been created to recognize diverse terrains. Through the integration of this CNN, an Android application is developed for identifying different types of terrain. This pioneering functionality effectively addresses a significant deficiency in current systems, delivering vital insights to VI users regarding the specific characteristics of the ground they are traversing.
- Hazard Detection: The system's ability to detect road puddles or wet floors is a key contribution that sets it apart from conventional intelligent white cane systems. By proactively alerting users to potential hazards, the approach reduces the risk of slips, falls, and accidents, ensuring a safer travel experience for individuals with low vision, especially the elderly.
- Developed a dataset to train and test the proposed model: A dataset is made by combining standard datasets like the Stagnant water dataset, and the GTOS-Mobile dataset, collecting images from Google and images captured using a mobile phone.
- Real-time Audio Feedback: To promote effective communication between the system and the user, the integration of audio feedback proves invaluable. By conveying information through auditory cues, the system ensures that VI individuals receive timely updates

\*Corresponding authors.

about the terrain and detected hazards, allowing them to make informed decisions promptly.

In summary, our approach introduces a novel lightweight CNN with the ability to identify different terrains. We curate a specialized dataset for training and model development. Employing this innovative CNN, an Android application is developed that offers terrain insights and alerts users through audio feedback, particularly notifying them about possible hazards such as slippery floors or puddles. This approach prioritizes human needs, safety, and independence, with the potential to revolutionize assistive technology and significantly enhance the lives of individuals with limited vision, particularly elderly users, leading to an overall improved quality of life.

The structure of the paper is as follows: In Section II, we delve into the most recent advancements in intelligent tools designed [9] proposed a system explicitly for individuals with visual impairments. Section III provides a comprehensive explanation of our proposed methodology. We outline the datasets utilized in this study in Section IV. The experimental validation and outcomes of our proposed system are detailed in Section V. Finally, Section VI [9] proposed a system serves as the conclusion of this paper.

## II. RELATED WORKS

Vision forms one of the most important sense organs, which gives vital information about surroundings. Our sense of sight is responsible for 80 percent of what we perceive. Vision enables humans to interact with their surroundings. One of the side effects of vision loss is a lack of confidence in one's ability to travel safely. The placement of tactile ground surface indicators, unsafe sidewalks, and the presence of barriers on sidewalks are key challenges in comprehending the outdoor difficulties experienced by VI individuals [4]. This section reviews some of the critical works in the field of guiding aids for VI people. This analysis is split into related works regarding existing smart white canes and terrain recognition.

### A. Robotic White Cane

The white cane is the most frequently used aid by individuals with visual impairments. Its main function is to help users assess their environment for possible dangers. Additionally, it aids in signaling to others that the user is blind, ensuring they receive appropriate assistance. In recent years, numerous enhanced versions of the white cane have emerged, offering a range of capabilities. Many of these innovations concentrate on tasks such as avoiding obstacles, identifying objects, and facilitating outdoor navigation.

Anwar et al. [1], introduces a smart cane equipped with an alarm to aid individuals with visual impairments when navigating challenging paths. An RF remote transmitter and receiver make it a unique electronic stick. It is equipped with an ultrasonic sensor and a buzzer. The VI person's walking path is detected using an ultrasonic sensor. At the same time, a global positioning system (GPS) paired with a voice stick for navigation, allows users to learn their present location and distance from their destination via voice commands. However, this system is constrained to the local vicinity only. [5] proposed a smart walker navigation system that assists

VI individuals with difficulty walking. It is utilized for two purposes: local obstacle detection in the spatial information setting, and guided navigation for achieving the desired goal. Vibrotactile signals provide obstacle information or navigation commands to the user.

A voice-activated electronic stick whose goal is to give users the confidence to move around in new situations was introduced by [6]. This enhanced model comprises global system for mobile communications (GSM), GPS, and Ultrasonic technology. It also employs biological authentication and incorporates an emergency trigger in the alarm system. [7] proposed an enhanced white cane that aids the blind community in guiding by providing results for all 270 degrees from the smart security walking stick's position. Ultrasonic sensors with a wide beam angle assist in a variety of obstacle detection applications. It performs well at identifying obstructions in the user's path within a three-meter range. This technology is low-cost, lightweight, and energy-efficient, with a noticeable quick response time.

The traditional VI aiding technologies addressed the obstacle avoidance problem alone. Information about the terrain is also vital for safe and smooth navigation for the VI person. A generative adversarial network (GAN) model named DiscoGAN is created by [8] to effectively translate ground images into a tactile signal that can be presented by an off-the-shelf vibration device. In this way, it can provide a global perception of the surroundings. The research [9] proposed a system based on DeepLabv3+ that is an improved semantic segmentation network. The technology can be used on a mobile phone to assist VI people in both indoor and outdoor settings. The training dataset is preprocessed with an illumination-invariant transformation to reduce the impact of variations in light. Bashiri et al. introduced a novel framework in their study [10], employing deep neural networks to facilitate indoor navigation for visually impaired individuals.

Existing literature for puddle detection mainly uses object detection models. Some of the existing smart white canes include puddle or water detection using a moisture sensor. But if the battery is not charged, this will not work which is the main disadvantage faced by these augmented canes [11].

### B. Ground Terrain Recognition

The quality of the terrain dataset, the feature extraction method, and the classification algorithm for terrain features are the key factors that affect the performance of a terrain image recognition system [12]. These aspects are complementary, necessitating optimization and control throughout, to improve the terrain classification and generalization capacity of a new algorithm. Moreover, there exists a lack of publicly available terrain classification datasets for research purposes. The literature related to terrain recognition explained here used proprietary datasets in their study.

Terrain identification is critical for outdoor mobile robot gait planning, speed control, and observation of the surroundings, among other things. The study [13] opted for a system in which an outdoor mobile robot is built that runs on a terrain dataset and extracts the high-level elements of the terrain image from MobileNet and DenseNet using the migrating learning approach. Extent-of-Texture information proposed by

[14] is another unique way to ground-terrain recognition, which employs a CNN backbone feature extractor network to extract relevant information from ground terrain images and to model the extent of texture and shape information locally. The research [15] proposed a differential angular imaging Network (DAIN), where small angular variations in image capture provide an enhanced appearance representation that significantly improves recognition. The features of materials recorded in the angular and spatial dimensions are encoded in this innovative network architecture. A Deep Encoding Pooling network [16] is another method for ground terrain categorization. Semantic scene comprehension is critical for gaining a better grasp of the surroundings. The study [17] proposed a system based on a combination of the efficient residual factorized network (ERFNet) and the 3D point cloud with a pyramid scene parsing network (PSPNet). The above-mentioned models are complex models for which accuracy can be further improved. Furthermore, most of them have not been implemented in real time. Recently, Intelligent Vehicles (IV) methodologies have been applied to the development of navigation assistive systems for VI people.

The analysis of existing literature highlights significant research gaps in the field. Most robotic white canes prioritize addressing obstacles, overlooking the potential of offering a comprehensive understanding of the environment, especially in unfamiliar routes. Presently, there's a lack of systems that can effectively notify users about hazards like puddles, wet and slippery floors. Additionally, the accuracy of current ground floor classification systems could be enhanced. The primary objective of our proposed system is to furnish users with real-time information about outdoor ground conditions, particularly alerting them to road puddles and wet floors, through a custom CNN model that accurately predicts terrain. This crucial information is conveyed audibly to enhance user safety and experience.

### III. METHODOLOGY

The prevailing terrain recognition models predominantly feature intricate deep-learning architectures. Nonetheless, for individuals with visual impairments, conveying terrain details through an Android app holds more promise. This research strives to forge a path towards a simplified and computationally streamlined deep-learning model tailored for terrain recognition systems. To realize this vision, a direct approach is adopted, entailing the deployment of a specialized CNN model. Additionally, an Android application is meticulously crafted to enhance accessibility for VI individuals, facilitating seamless access to crucial terrain information.

A custom dataset consisting of five image classes, namely cement, grass, road, wet floor, and puddles has been created to help VI people better understand their terrain and be alerted of any potential hindrances. This dataset is a combination of GTOS-Mobile (Ground Terrain Outdoor Services) [18] data, Google images, IEEE Stagnant Water dataset, and ground terrain images captured using a mobile phone. An Android app has been developed which uses a custom-built CNN to detect ground terrain and provides audio feedback to the user with directions based on the detected features in real-time.

The proposed pipeline consists of three main steps, namely data pre-processing, CNN model design, and Android app

development. The images obtained from various data sets are resized to the same dimension in the pre-processing phase and are then augmented before feeding to the custom CNN. The pre-processing steps used were resizing, and gray scaling. The data augmentation involves random flip, random zoom, and random rotation. After model training, test data prediction and evaluation are done using the trained deep learning model. Fig. 1 shows the flow diagram of the proposed system for ground terrain detection. It represents the schematic of the proposed system.

#### A. Custom CNN

The Custom CNN architecture is comprised of five convolutional layers as depicted in Fig. 2. Detailed insights into the layer structure and parameter specifics can be found in Table I. The initial convolution layer block consists of two sets of convolution layers, each comprising 16 filters. The subsequent block contains two sets of convolution layers with 32 filters each, followed by a third block with a convolution layer featuring 64 filters. These layers employ the same padding scheme, along with ReLU activation and Max pooling. The filter size utilized in each of these convolutional layers is 3x3.

After these convolutional layers, a flattening layer is introduced, followed by a fully connected layer consisting of 128 filters, employing ReLU activation. Ultimately, a dense layer is incorporated, with the number of classes serving as a parameter. In summary, the Custom CNN model includes 560,000 trainable parameters, featuring five convolutional layers with associated Maxpooling layers and ReLU activation. It further incorporates a fully connected layer utilizing a dropout rate of 50%, followed by an output dense layer with the number of classes as a parameter. The model employs the sparse categorical cross-entropy loss function, particularly useful for multi-class image classification tasks. This loss function quantifies the cross-entropy loss between predictions and labels. Importantly, its utilization of integer representations for labels instead of vector representations contributes to efficiency in memory and computation. The mathematical formulation of this entropy function is:

$$L_{CE} = - \sum_{i=1}^n t_i \log(p_i)$$

where  $t_i$  is the truth label and  $p_i$  is the Softmax probability for the  $i^{\text{th}}$  class and  $n$  represents the number of classes.

The architecture of the Custom CNN is depicted in Fig. 2. To make the test results more accurate, we fine-tuned the CNN model so that it could better understand the test data. By doing this, the model became better at recognizing different features in the test images. This process involves making small improvements to how the model learns from the given data. The combination of the CNN's design, how it extracts features, and the fine-tuning process all work together to create a strong model that can accurately identify and classify images. The images were resized to dimensions of  $64 \times 64$  pixels and a batch size of 15 is selected empirically. The number of filters within the fully connected Layer is fixed as 128.

### IV. DATA COLLECTION

We curated an exclusive dataset tailored to our proposed system. This dataset comprises diverse visual representations

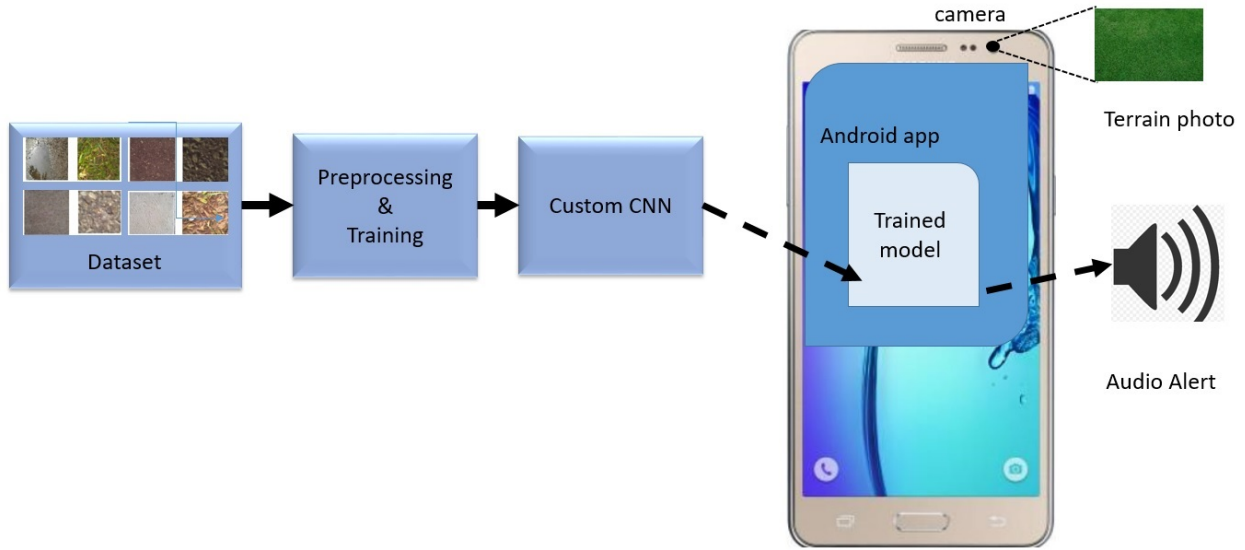


Fig. 1. Flow diagram of the proposed system for ground terrain detection.

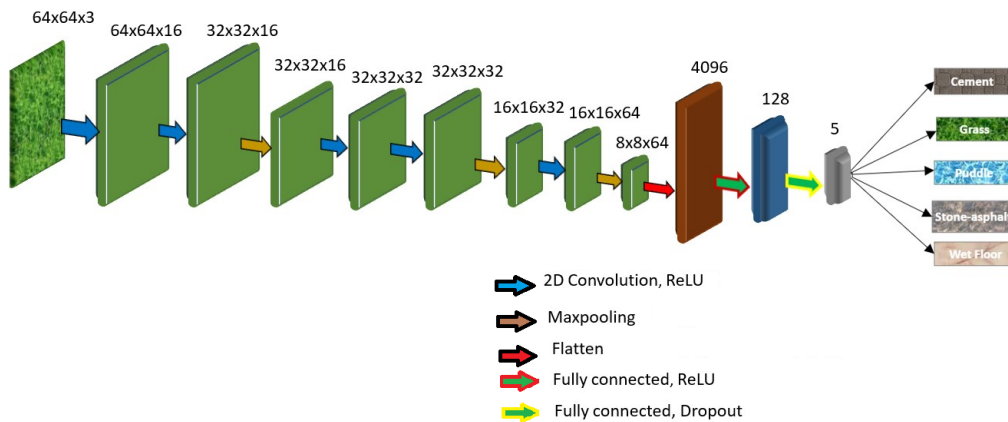


Fig. 2. Architecture of the proposed custom CNN for identifying ground terrain.

TABLE I. LAYER STRUCTURE DETAILS OF MODIFIED CUSTOM CNN

Layer	Input	Kernels	Output	Parameters
Conv1	$64 \times 64 \times 3$	16	$64 \times 64 \times 16$	448
Conv2	$64 \times 64 \times 16$	16	$64 \times 64 \times 16$	2320
Maxpool1	$64 \times 64 \times 16$	-	$32 \times 32 \times 16$	-
Conv3	$32 \times 32 \times 16$	32	$32 \times 32 \times 32$	4640
Conv4	$32 \times 32 \times 32$	32	$32 \times 32 \times 32$	9248
Maxpool2	$32 \times 32 \times 32$	-	$16 \times 16 \times 32$	-
Conv5	$16 \times 16 \times 32$	64	$16 \times 16 \times 64$	18496
Maxpool3	$16 \times 16 \times 64$	-	$8 \times 8 \times 64$	-
Flatten	$8 \times 8 \times 64$	-	4096	-
Dense1	4096	128	128	524416
Dense2	128	-	5	645

of outdoor ground surfaces, such as concrete, grass, and asphalt-covered stone. Additionally, it incorporates imagery depicting wet patches and puddles. Our dataset draw images from the GTOS-Mobile dataset [18] for ground terrain information, while images featuring puddles and damp floors were

sourced from the IEEE Stagnant Water dataset [19]. Furthermore, we enriched the dataset with video footage recorded using a mobile device.

#### A. GTOS-Mobile Dataset

The GTOS-Mobile dataset [18] encompasses 81 videos that showcase terrain categories similar to those found in the GTOS dataset. These videos were captured using a handheld mobile phone, introducing a variety of lighting conditions and viewpoints. The dataset comprises an extensive collection of 100,000 images distributed across 31 distinct classes. After a meticulous curation process we extracted 6066 frames by employing a temporal sampling rate of approximately 1/10th of a second. However, the emphasis is placed solely on the crucial ground terrain categories. Owing to concerns surrounding image clarity, data pertaining to sand and soil from the GTOS-Mobile dataset is deemed less dependable, prompting its exclusion from consideration. Consequently, for the purpose of outdoor terrain recognition, the analysis narrows

down to the inclusion of the cement, grass, and stone-asphalt classes extracted from this dataset. Fig. 3 shows sample images extracted from this data set.

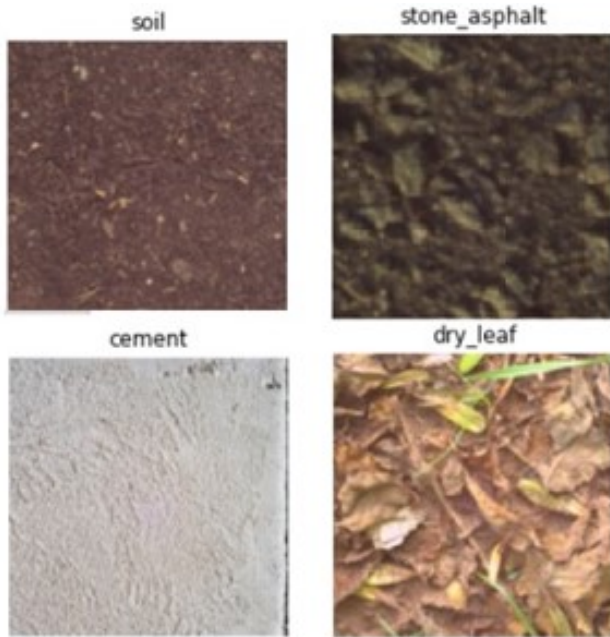


Fig. 3. Images from GTOS-Mobile dataset.

### B. Stagnant Water Dataset

The dataset [19] comprises images with dimensions of  $256 \times 256$  pixels captured using a mobile device. This dataset comprises a total of 1976 annotated photos and encompasses two categories, namely moist surfaces and bodies of water. The dataset is further divided into distinct categories: Indoor, Outdoor, and Raw data. The raw data category comprises unprocessed images, each accompanied by a specified range of resolutions. Puddle detection presents a range of complexities due to fluctuations in lighting conditions, diverse image capture angles, and the presence of reflections on puddle surfaces. To tackle these challenges, approaches such as data augmentation and the collection of images from a diverse range of angles and lighting conditions have been implemented. Fig. 4 shows sample images from the Stagnant Water dataset.

### C. Custom Dataset

A unique dataset was compiled by capturing images using a Redmi Note 7 mobile phone with a 12-megapixel camera. This dataset encompasses 150 images for each distinct category: cement, grass, stone-asphalt, puddle, and wet floor. Images taken by VI persons using mobile phone might not exhibit flawless quality. Intentionally, the custom dataset includes both shaky and low-quality images, mimicking real-world conditions. This deliberate inclusion aims to provide the training process with a more realistic representation of the environment. The images from this custom dataset are depicted in Fig. 5.

The final dataset contains 16763 files belonging to five classes. The images are acquired from the GTOS-Mobile dataset, Stagnant water dataset and images captured using our



Fig. 4. Images from stagnant water dataset.



Fig. 5. Images from custom dataset.

own mobile device. The training data is split into train data and validation data with a validation split 20%. A total of 12973 files is used as training set, 3243 files for validation, and 759 files as test dataset. Table II shows number of images in each class of custom dataset.

## V. RESULTS AND DISCUSSION

The CNN model proposed for terrain classification underwent a comprehensive training process to realize the required customized model for accurate terrain detection. The training utilized a specialized dataset encompassing a variety of outdoor terrains, including cement, stone-asphalt, grass, as well as distinct classes for puddles and wet floors. In the

TABLE II. NUMBER OF IMAGES IN EACH CLASS OF CUSTOM DATASET

Class	Train dataset images	Test dataset images
Cement	7550	174
Grass	3100	196
Puddle	1659	114
Stone-asphalt	2370	154
Wetfloor	2084	121

initial stages of training, validation accuracy surpassed 95%, while testing accuracy remained below 70%, indicative of an overfitting scenario. To address this challenge, a combination of techniques including early stopping, dropout, and data augmentation was employed, resulting in a notable reduction of this accuracy gap.

Efforts were directed towards enhancing testing accuracy through layer-wise modifications in the CNN architecture. These refinements led to a remarkable increase in testing accuracy to 98%. The Adam optimizer, with a default learning rate of 0.001, was leveraged to optimize the modified Custom CNN model. Training the model for 300 epochs with Early Stopping set at 50 epochs based on minimum validation loss yielded promising outcomes.

The accuracy and loss curves of the Custom CNN model are visually depicted in Fig. 6, illustrating the progressive improvement achieved during the training process. Meanwhile, Fig. 7 presents the confusion matrix, offering insights into the model’s performance across various terrain classes. It is important to note that the observed oscillations in the curves can be attributed to the utilization of an imbalanced dataset in this study. Future work could involve mitigating such oscillations by considering a slightly lower learning rate or implementing exponential decay learning rate strategies. The learning rate, a pivotal parameter in optimization, requires a balanced selection that ensures a trade-off between convergence speed and overshooting tendencies. As evidenced by validation loss values ranging from 0.1 to 0.4, finding this optimal balance remains a crucial consideration.

Further advancements were achieved by enriching the puddle class with additional images sourced from the Stagnant Water dataset. This augmentation elevated the model’s validation accuracy to an impressive 98%, while testing accuracy reached 94%, as depicted in Fig. 8. The confusion matrix revealed that, post-enhancement, instances of misclassification emerged wherein puddles were sometimes classified as wet floors. It is worth noting that road puddles and wet floors both fall under the puddle category. The mispredictions of this nature do not significantly impact system functionality, as the system’s role is to provide audio feedback advising users to proceed cautiously in case of detected puddles or wet floors.

The proposed CNN model demonstrates the remarkable potential for real-time terrain classification. The results highlight avenues for further refinement, including domain transfer implementations. The integration of this model into practical applications, particularly those necessitating immediate feedback for user safety, underscores its relevance and efficacy in

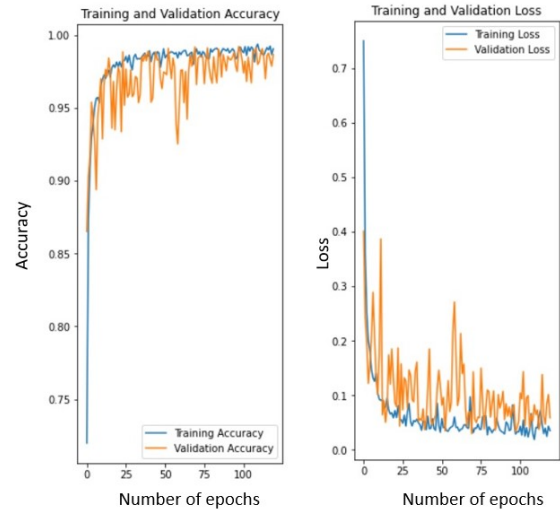


Fig. 6. Accuracy and loss curves of the custom CNN model.

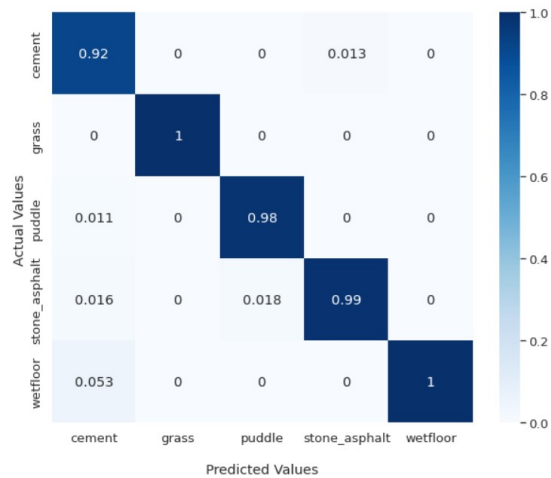


Fig. 7. Confusion matrix of the custom CNN model.

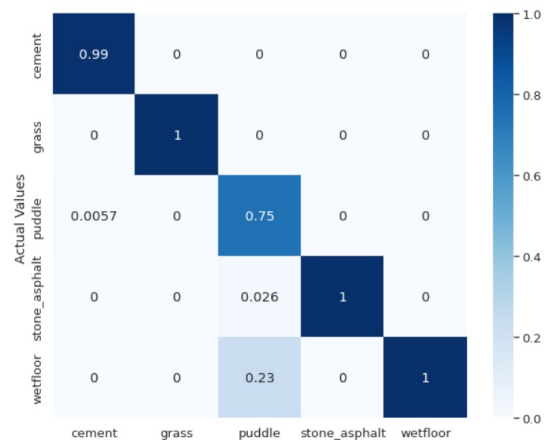


Fig. 8. Confusion matrix of the model after enhancing the puddle dataset.

real-world scenarios.

### A. Android Application Results

The central objective behind the proposed system is to enhance the capabilities of conventional robotic white canes by introducing a comprehensive understanding of the surrounding environment. This innovation aims to empower individuals with visual impairments (VIs) by seamlessly integrating global insights into their daily navigation. Rather than augmenting the white cane with additional hardware and algorithmic components, a more user-friendly approach involves encapsulating this functionality within an Android application.

To achieve this goal, a fundamental Android application was developed as a preliminary step to assess the real-time behavior of the proposed deep learning model. This undertaking laid the groundwork for subsequent domain transfer implementations. The transition from a deep learning model stored in the .h5 format to a .tflite model was a pivotal phase. The choice between a quantized or unquantized model depended on size considerations. The deep learning model, which clocked in at 2MB in its original form, maintained this size when converted into an unquantized .tflite model. Opting for a quantized eight-bit .tflite model reduced the size to 500KB. However, this reduction in size through quantization came at a trade-off with accuracy.

The development of the Android application was facilitated by the widely used Android Studio framework, which supports the Java programming language and is well-suited for mobile device applications. After a thorough simulation process, the deep learning model with the lowest validation loss was saved using model checkpointing. To ensure optimal performance on mobile devices, this selected model was converted to an unquantized .tflite model. By doing so, the .tflite model retained a manageable size of 2MB, enabling its seamless integration into the Android application.

The proposed system's core concept revolves around extending the functionality of robotic white canes through a user-friendly Android application. By encapsulating complex insights into a lightweight .tflite model, individuals with visual impairments can gain an enhanced understanding of their surroundings, thereby advancing their autonomy and safety in navigation.

The initial phase of development involved creating a basic version of the Android App, which serves as a foundation for evaluating the system's real-time performance. This App captures images and employs the deep learning model to predict the type of terrain and identify the presence of puddles or wet floors. The primary objective of this basic App version is to gain insights into the system's behavior under real-world conditions.

For testing purposes, the App's user interface includes two buttons, as illustrated in Fig. 9. One button captures an image in real time, while the other allows users to select an image from their gallery. These preliminary App results serve as a foundation for domain transfer evaluations. It is important to emphasize that the basic version of the app is not tailored for individuals with visual impairments (VIs). To address this, an advanced Android app was developed to provide continuous

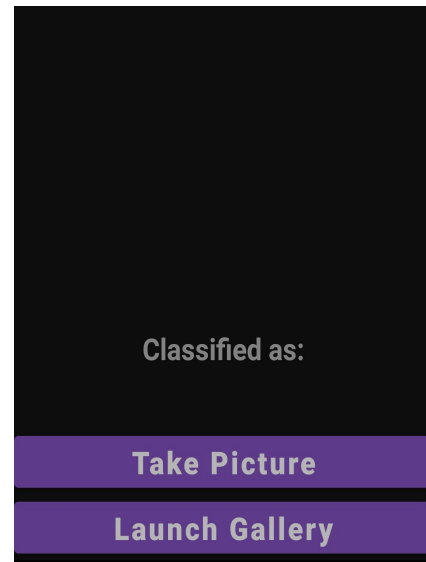


Fig. 9. Android application Interface. Image shows screenshot of the android app, created for testing purposes.

video monitoring and audio feedback, ensuring its suitability and effectiveness for VI users.

## VI. CONCLUSIONS

In recent decades, advancements in technology have led to the development of robotic and augmented white canes tailored for individuals with visual impairments (VIs). These innovations primarily focus on object detection, obstacle avoidance, and navigation. The major challenge involves providing VI individuals with a comprehensive understanding of their environment to facilitate improved interaction with their surroundings during travel. The central aim of the proposed system is to furnish VI individuals with an encompassing perception of their surroundings. This is achieved through the identification of diverse ground terrains, encompassing cement, grass, and asphalt, along with the detection of road puddles and wet floors. To realize this, a Custom CNN is employed as a deep learning model for a multi-class image classification challenge. The system culminates in conveying vital information to users through audio feedback.

Following an intensive training regimen, the deep learning model achieves commendable performance, boasting a testing accuracy of 94% and an impressive validation accuracy of 98%. To facilitate domain transfer, an Android app was meticulously designed, enabling real-time testing to assess the robustness of the proposed system in various scenarios.

The future trajectory of the proposed system holds significant promise. Plans encompass the expansion of more ground terrain classes to broaden the system's scope and versatility. Additionally, there are intentions to tailor the Android app for individuals with low vision impairments, incorporating continuous video monitoring. This enhancement ensures accurate terrain and puddle identification, even under varying illumination conditions. The App's connection with users will be fortified through intuitive audio feedback, reinforcing its accessibility and usability.

REFERENCES

- [1] A. Anwar and S. Aljhdali, "A smart stick for assisting blind people," *IOSR Journal of Computer Engineering*, vol. 19, no. 3, pp. 86–90, 2017.
- [2] X. Wang, J. Calderon, N. Khoshavi, and L. G. Jaimes, "Path and floor detection in outdoor environments for fall prevention of the visually impaired population," in *2022 IEEE 19th Annual Consumer Communications & Networking Conference (CCNC)*. IEEE, 2022, pp. 1–6.
- [3] H. Walle, C. De Runz, B. Serres, and G. Venturini, "A survey on recent advances in ai and vision-based methods for helping and guiding visually impaired people," *Applied Sciences*, vol. 12, no. 5, p. 2308, 2022.
- [4] A. Riazi, F. Riazi, R. Yoosfi, and F. Bahmeci, "Outdoor difficulties experienced by a group of visually impaired iranian people," *Journal of current ophthalmology*, vol. 28, no. 2, pp. 85–90, 2016.
- [5] A. Wachaja, P. Agarwal, M. Zink, M. R. Adame, K. Möller, and W. Burgard, "Navigating blind people with walking impairments using a smart walker," *Autonomous Robots*, vol. 41, pp. 555–573, 2017.
- [6] R. S. Kolhe, K. G. Dhole, P. S. Thakre, P. S. Prasad, P. S. Patel, A. V. Rodge, and F. Akhtar, "Smart stick for the blind and visually impaired people," 2021.
- [7] P. Mind, G. Palkar, A. Mahamuni, and S. Sahare, "Smart stick for visually impaired," *Int. J. Eng. Res. Technol*, vol. 10, no. 06, pp. 196–198, 2021.
- [8] H. Liu, D. Guo, X. Zhang, W. Zhu, B. Fang, and F. Sun, "Toward image-to-tactile cross-modal perception for visually impaired people," *IEEE Transactions on Automation Science and Engineering*, vol. 18, no. 2, pp. 521–529, 2020.
- [9] C. Xu, K. Wang, K. Yang, R. Cheng, and J. Bai, "Semantic scene understanding on mobile device with illumination invariance for the visually impaired," in *Artificial Intelligence and Machine Learning in Defense Applications*, vol. 11169. SPIE, 2019, pp. 218–226.
- [10] F. S. Bashiri, E. LaRose, J. C. Badger, R. M. D'Souza, Z. Yu, and P. Peissig, "Object detection to assist visually impaired people: A deep neural network adventure," in *Advances in Visual Computing: 13th International Symposium, ISVC 2018, Las Vegas, NV, USA, November 19–21, 2018, Proceedings 13*. Springer, 2018, pp. 500–510.
- [11] A. A. Elsonbaty, "Smart blind stick design and implementation," *International Journal of Engineering and Advanced Technology (IJEAT)*, vol. 10, no. 5, 2021.
- [12] P. Kozłowski and K. Walas, "Deep neural networks for terrain recognition task," in *2018 Baltic URSI Symposium (URSI)*. IEEE, 2018, pp. 283–286.
- [13] S. Zeng, H. Huang, and Z. Shi, "Outdoor terrain recognition based on transfer learning," in *Journal of Physics: Conference Series*, vol. 1846, no. 1. IOP Publishing, 2021, p. 012012.
- [14] S. Ghose, P. N. Chowdhury, P. P. Roy, and U. Pal, "Modeling extent-of-texture information for ground terrain recognition," in *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, 2021, pp. 4766–4773.
- [15] J. Xue, H. Zhang, K. Dana, and K. Nishino, "Differential angular imaging for material recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 764–773.
- [16] J. Xue, H. Zhang, and K. Dana, "Deep texture manifold for ground terrain recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 558–567.
- [17] K. Yang, L. M. Bergasa, E. Romera, R. Cheng, T. Chen, and K. Wang, "Unifying terrain awareness through real-time semantic segmentation," in *2018 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2018, pp. 1033–1038.
- [18] J. Xue, H. Zhang, and K. Dana, "Deep texture manifold for ground terrain recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 558–567.
- [19] S. Bhutad and K. Patil, "Dataset of stagnant water and wet surface label images for detection," *Data in Brief*, vol. 40, p. 107752, 2022.



# ArCyb: A Robust Machine-Learning Model for Arabic Cyberbullying Tweets in Saudi Arabia

Khalid T. Mursi<sup>1</sup>, Abdulrahman Y. Almalki<sup>2</sup>,

Moayad M. Alshangiti<sup>3</sup>, Faisal S. Alsubaei<sup>4</sup>, Ahmed A. Alghamdi<sup>5</sup>

Department of Cybersecurity, University of Jeddah, Jeddah, Saudi Arabia<sup>1,2,4,5</sup>

Department of Software Engineering, University of Jeddah, Jeddah, Saudi Arabia<sup>3</sup>

**Abstract**—The widespread use of computers and smartphones has led to an increase in social media usage, where users can express their opinions freely. However, this freedom of expression can be misused for spreading abusive and bullying content online. To ensure a safe online environment, cybersecurity experts are continuously researching effective and intelligent ways to respond to such activities. In this work, we present ArCyb, a robust machine-learning model for detecting cyberbullying in social media using a manually labeled Arabic dataset. The model achieved 89% prediction accuracy, surpassing the state-of-the-art cyberbullying models. The results of this work can be utilized by social media platforms, government agencies, and internet service providers to detect and prevent the spread of bullying posts in social networks.

**Keywords**—Natural language processing; machine learning; neural network; bullying; cyberbullying

## I. INTRODUCTION

The widespread use of computers and smartphones has greatly increased the use of social media in recent years. Social media platforms allow users to express their opinions and emotions freely, either using their real identities or anonymously. Unfortunately, this freedom of expression has also led to the spread of online bullying. People can hide behind anonymity to harass and bully others, causing significant harm and distress to their victims. It's important for social media companies and individuals to take steps to prevent and address this issue and to ensure that these platforms remain safe and positive spaces for everyone.

Cyberbullying refers to any deliberate aggressive behavior via social media done by an individual or a group of individuals that post offensive or hostile messages that result in discomfort or harm to other users [1]. Dani et al. in [2] defined cyberbullying as the phenomena of intentionally harassing or abusing others through cell phones, internet, and other electronic devices. According to [3], cyberbullying is confirmed as a serious global problem that should be confronted and prevented from spreading. Cyberbullying is worse and more insidious than traditional bullying and has severe consequences since it is not restricted to a time or a place. The bullying content can be posted in a single action as a comment or a tweet by an abuser. Cyberbullying enables the perpetrator with the ability to humiliate or embarrass the victim in plain sight. Also, this content can be viewed, saved, shared, quoted, or liked by others multiple times, resulting in an ongoing cycle of the original assault creating persistent damage or distress for the victim. Cyberbullying victims suffer from depression,

anxiety, low self-esteem, anger, frustration, feelings of fear, and in tragic scenarios the victims attempt suicide [4]–[6].

Moreover, in [2], the author stated that cyberbullying is becoming more frequent due to the growth of social media platforms. A study was done by Al-Zahrani [3] investigating cyberbullying in Saudi Arabia among higher-education students, 287 students participated in the study, as 26.5% of the students admitted that they have cyberbullied others at least once, while the majority of students 57% have witnessed cyberbullying once or twice on at least one student. He concluded that cyberbullying rate in Saudi Arabia has increased by 9% during the study time period.

Twitter is a microblogging platform that allows users to express their opinions and share their thoughts. Users can follow influencers, brands, and news accounts to stay informed about current events and trends [7]. As noted by Alasem [8], the number of Twitter accounts in Saudi Arabia has been increasing, with the platform experiencing fast growth in the country. In 2012, Riyadh, the capital city of Saudi Arabia, was ranked as the tenth most active city globally in terms of statistics and tweets. Given the vast array of topics and trends that emerge on social media, cyberbullying can take on various forms and can be challenging to identify.

Detecting cyberbullying on social media requires an understanding of users' opinions, tweets, and emotions, which can then be analyzed to determine whether the content constitutes bullying or not. According to Saberi and Saad [9], sentiment analysis involves the detection, extraction, and classification of opinions or comments on a particular topic. The primary goal of sentiment analysis is to classify the opinion, comment, or blog as either positive, negative, or neutral. However, detecting cyberbullying in the Arabic language presents significant challenges due to its complex structure, diverse dialects, informal language used on social media, and wide range of synonyms. Additionally, the Arabic language in social media is often written with diacritics that aid in pronunciation, making normalization, tokenization, and stemming difficult to apply.

To ensure effective detection of Arabic cyberbullying comments on social media, a well-trained machine-learning model is essential. This is particularly important given the widespread use of the Arabic language, which is spoken by approximately 420 million people [10]. However, developing such a model is challenging due to the lack of labeled Arabic datasets and research on this topic. As of the writing of this paper, there is no well-trained model with more than 90% prediction accuracy for detecting Arabic cyberbullying

comments. Furthermore, as detailed in Dani et al. [2], detecting and combating cyberbullying in the Arabic language presents several challenges, including the nature of online comments and reviews. These comments are often unstructured, short, and obfuscated, making it difficult to identify common patterns in machine-learning models. To address these challenges, we present the following in our work:

- A comprehensive analysis on Arabic cyberbullying tweets and their growth over time.
- A novel Arabic cyberbullying dataset labeled using a rigorous methodology.
- A deep learning model that can detect Arabic cyberbullying with a prediction accuracy that is equal to or better than the state of the art.

The rest of the paper is organized as follows: Section II presents the background information and preliminaries of the cyberbullying in the Arabic language. Section III presents our proposed deep learning detection method. In Section IV, we discuss the experimental setup for our experiments. In Section V, we discuss the experimental results obtained. Section VI concludes the paper.

## II. BACKGROUND

Many studies have been published in the sentiment analysis field. Researchers have provided interesting methods and approaches contributing to this field improvement. Abdul-Mageed et al. in [11] produced an Arabic dataset that was divided into four classes, objective, subjective-positive, subjective-negative, and subjective-neutral, and was manually labeled. The authors followed classification criteria that were taken from [12] in which if a phrase is not objective, it will fall into one of the three subjective classes. Out of their strict annotation process, their dataset consists of 1281 objective, 491 subjective-positive, 689 subjective-negative, and 394 subjective-neutral news sentences. Then for the classification, they've done two stages using the SVM classifier with linear kernel. The first stage for classifying the subjectivity, train the model to differentiate the subjective and objective sentences, and the second stage is to study the sentiment, and train the model to differentiate the positive and negative subjective sentences. As a result of their work, they obtained 65% and 52% *F*-score for the subjectivity and sentiment studies respectively.

Duwairi and Qarqaz in [13] used open-source software with a graphical user interface to build their machine learning model. They have generated a dataset from Twitter and Facebook that consist of 2591 tweet and comment, 1073 positive and 1518 negative samples, and were classified using a crowdsourcing tool. The dataset addresses multiple topics such as sports, education, and political news. The Naïve Bayes, KNN, and SVM classifiers were used, the SVM achieved a higher precision rate and it equals 75.25%.

Shoukry and Rafea in [14] have used machine learning to study the Arabic sentiments. They collected more than 4000 tweets and then finally have extracted 1000 tweets consisting of 500 negative and 500 positive tweets. Their tweet extraction targeted tweets that only hold one opinion and avoided sarcastic and subjective tweets. For the feature extracting

Shoukry and Rafea method was revealed from [15] where the statistical machine-learning is implemented to highlight the most common words to act as candidate features. For the classification task, they used the Weka software to classify the tweets using Naïve Bayes and SVM with accuracy around 65% and 72%, respectively.

Al-Kabi et al in [16] developed an analysis tool that can classify the opinions and comments based on standard and slang Arabic forms. One of the tool tasks is classifying the text into positive or negative, which is indirectly related to our research. In specific, their dataset consists of reviews collected from 72 social media websites with a total of 1080 reviews, and their machine-learning method was the Naïve Bayes. Their method successfully identified the subjectivity, polarity, and intensity of the Arabic reviews with prediction accuracy around 90%, 93%, and 96%, respectively.

AL-Rubaiee et al. in [17] implemented NLP and machine learning to classify tweets according to their sentiment polarity. Their work concentrated on opinion mining in a trading strategy with Mubasher products, a stock analysis software in Saudi Arabia, which made it considered topic-specific in the field of the sentiment analysis of the Arabic language. They collected and manually labeled around 1331 tweets by two experienced Mubasher employees. Therefore as a result of their annotation process, their dataset consists of 378 positive, 755 negative, and 198 neutral tweets. The prediction accuracy of their Naïve Bayes and SVM model are around 83% and 79% respectively.

On the other hand, There are a few other topic-specific works that focus on constructing machine-learning models to detect cyberbullying behavior. In 2019 AlHarbi et al. published the first work in the Arabic cyberbullying field [18]. In specific, they built a lexicon-based model that consist of more than 100K samples. They used R language for data extraction, 50K tweets, and 50K Youtube comments. They were able to obtain 81% prediction accuracy for the trained cyberbullying model. Similarly, Almutiry and Fattah in [19] collected a dataset automatically through Twitter API and ArabiTools with a total of 17748 tweets. they followed two collecting methodologies, one is query-oriented by searching for specific keywords, and the other is random selection. The dataset was labeled by both means manual and automatic. The automatic labeling was done by considering the nature of the tweet, if a tweet contains cyberbullying words it will be labeled as cyberbullying, and otherwise non-cyberbullying. After collecting and labeling the dataset a couple of steps were performed in preprocessing such as Normalization, Tokenization, ArabicStemmerKhoja, Light Stemmer, and Term Frequency-Inverse Document Frequency(TF-IDF). Then for the classification, they used the SVM algorithm with both Python and WEKA. After performing three experiments WEKA results showed the highest efficiency with 85.49% prediction accuracy. However, we argue that relying on automatic labeling is not currently practical. Given the significance of accurate sample annotations, utilizing automatic labeling techniques would introduce a substantial risk of duplicating samples and producing inaccurate classifications. Therefore, manual labeling remains indispensable until we develop a highly accurate model capable of consistently and reliably classifying the samples.

Almutairi and Alhagry in [20] started the data collection

process using Twitter API and collected a total of 8154. The authors focused on collecting their dataset from Saudi Arabia. The data collection process spanned approximately one year and seven months, capturing tweets related to various events such as student exams, vacations, and the COVID-19 pandemic. During the preprocessing phase, they applied several cleaning steps, including the removal of URLs, mentions, emojis, hashtags, newlines, repeated letters, digits, Arabic diacritics, and unrelated tweets. For the classification task, they employed multiple machine learning algorithms and found that the SVM algorithm achieved the highest prediction accuracy of 82

As shown in the literature above and besides some works that were not mentioned [21]–[24], in the past ten years, the Arabic sentiment analysis got a lot of attention in several topics such as users reviews in the trade market, positive and negative tweets, and cyberbullying. Nevertheless, there is neither a publicly available Arabic cyberbullying dataset nor a well-trained machine learning model for cyberbullying due to the difficulty of the Arabic language and its many dialects, along with the slang language used by the majority of Arab users. Therefore, our work contributes to the research community by providing a well-trained machine learning model based on a manually labeled dataset.

### III. METHODOLOGY

This section presents the methodology used to build a machine-learning model for Arabic sentiment analysis and cyberbullying detection. The process begins with a discussion of the data collection method and labeling process. Next, the proposed preprocessing and machine learning approach are presented. Finally, the evaluation of the approach is discussed. To provide an overview of the approach and steps taken to detect cyberbullying in social platforms, Fig. 1 is presented.

#### A. Data Collection

As discussed in Section II, the lack of publicly available datasets on cyberbullying in the Arabic language posed a significant challenge for this study. Hence, the collection and labeling of a suitable dataset proved to be a time-consuming and challenging task, presenting the most significant hurdle in the project. To kickstart the project, we formed multiple teams and manually analyzed the Twitter space to familiarize ourselves with the terminologies and behaviors associated with cyberbullying on social media, as well as the techniques used by perpetrators. This led us to identify 16 keywords, including *khibel*, *Abd*, and *Marid*, which we used to collect the dataset. Table I presents some of the search keywords that we employed in this research. We hand-selected these terms based on our examination of the most prevalent Arabic bullying terms. We believe that most tweets utilizing these terms are likely to be of a bullying nature. Once the search keywords were determined, we utilized the Twitter-API, which is publicly accessible, to retrieve tweets containing the designated keywords. Each downloaded tweet was required to contain at least one of the specified keywords. During the data collection phase, we encountered various obstacles, such as the maximum number of allowed tweets to collect per day by the API, tweets written in foreign dialects and languages, and a high number of duplicate tweets. However, these obstacles are not unique to

our study and are commonly encountered in similar research. To overcome these challenges, we adopted best practices and strategies from previous studies.

TABLE I. EXAMPLES OF CYBERBULLYING KEYWORDS USED IN SOCIAL PLATFORMS AND THEIR ENGLISH TRANSLATION

Keyword	English meaning	Arabic Keyword
Khibel	Person that lacks intelligence	خبيل
Tays	Person that lacks common sense	تيس
Abd	Slave	عبد
Marid	Pervert or Sick	مريض
Nafsiyah	Refers to the person's psychological state	نفسية
Immah	Person who blindly agrees with someone regardless of their actions	امعه
Baka	Whiner	بكي
Madala	Spoiled	مدلع
Moaq	Handicapped (insult)	معاق
Bahima	Animal (insult)	بهيمه
Kalb	Dog (insult)	كلب
Maafen	Disgusting	معفن
Ghabi	Stupid	غبّي
Wajhk	Your face	وجهك
Seyah	Screaming in literal meaning but could also mean crying or whining	صياح
Yifashil	Embarrassing	يفشل

#### B. Labeling

Our dataset comprises 4140 samples, of which 2070 tweets are labeled as bullying and 2070 tweets are labeled as non-bullying. We decided to remove the user's identities to protect their privacy. For our cyberbullying tweets, manual labeling is necessary due to the absence of diacritics in written Arabic, which represent vowels. The lack of diacritics generates ambiguity, which increases the range of possible interpretations in the Arabic language [25]. Additionally, bullying can be disguised in a normal sentence that cannot be detected by automated labeling tools. Every tweet in the dataset was independently annotated by two cybersecurity specialists, all of whom are native Arabic speakers. The annotation process took place over a period of six weeks. In cases where conflicts arose among the annotators, a third specialist was involved to resolve them through discussions with the two cybersecurity specialists.

Table II lists some samples that were difficult to classify because of their confusing nature. For example, the first sample states "Finally it's my favorite time where I go to sleep and put my phone on silent while others whine while they go to work/school". One can build a case that this is indirect bullying to those who need to go to work from those with the luxury to stay home. However, you can also build a case that the user is describing their feelings without interfering with or offending anyone. In our case, we followed the later logic since the user did not use any offensive language, which is usually included based on the commonly accepted definition for bullying [1] [2]. In the second example, a user responds to a tweet announcing that schools' final exams will be held on campus, which has sparked complaints from students who have been attending

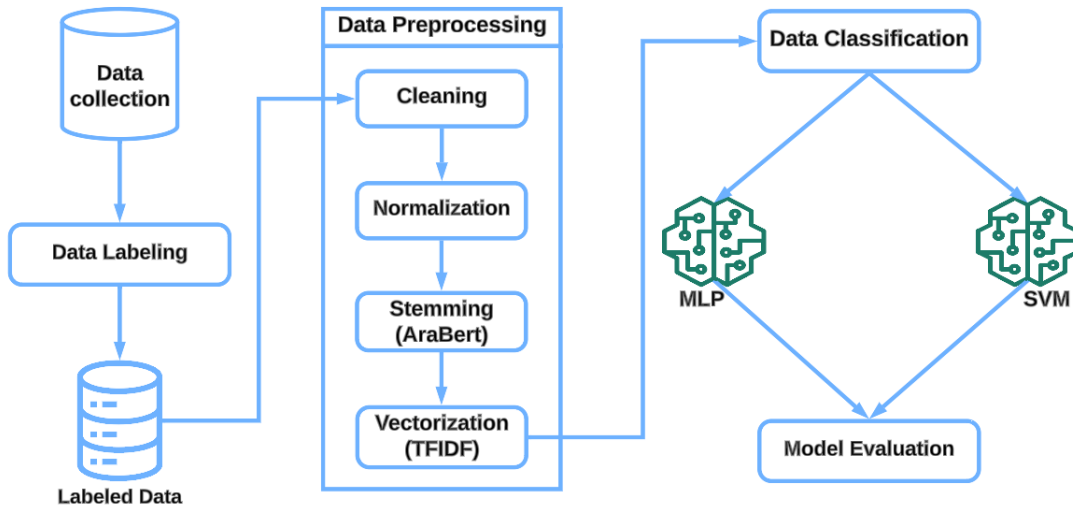


Fig. 1. The research methodology. The figure illustrates the sequential steps undertaken in this study, starting with the collection of raw data, then the labeling of the data, followed by preprocessing, modeling using AraBERT and TFIDF, and concluding with the analysis of the model’s result. The arrows indicate the flow and progression of the research process.

school online due to the COVID-19 lockdowns. The user says "They deserve it, let them test on campus. They disgust me. They want to stay at home and in bed forever. What a spoiled and useless generation". The user’s response contains offensive words and implies that the current generation is lazy, soft, and unsuccessful. You may build a case that the user is not targeting someone specifically, and you also may build a case that he is targeting a large but specific set of people. We believe the user identified the entity they are offending and bullying and used multiple offensive words in the tweet, so we classified this tweet as an instance of cyberbullying, following the definition proposed by [1], [2].

TABLE II. EXAMPLES OF TWEETS THAT ARE DIFFICULT TO CLASSIFY

Tweets	Label
اخيرا جا وقت ففرقي المفضلة تصميت الجوال والنوم على انغام صياح المداومين	0
يستاهلون نبيهم يختبرون حضوري قرفونا لين متى الدلع والتسدرح جيل مدلع و فاشل	1

Table III lists some samples from our dataset. Those examples also show that the same word can have completely different meanings depending on the context in which it is used. In our example, we focused on the keywords Nafsiyah and Khibel. Let’s first examine the samples that use the keyword Nafsiyah. The first tweet in Table III, which we labeled as non-bullying, uses the keyword in a positive manner where the user expresses relief for completing a month without going to any health clinics. However, in the second tweet, which we classified as a bullying tweet, the user replied with an accusation that the original tweet author has psychological issues. Now, let’s examine the samples that use the keyword Khibel. The third tweet in Table III, which we classified as a non-bullying tweet, uses the term to describe someone with a humorous and entertaining personality. On the other hand, in the fourth tweet, the user replies to another tweet,

criticizing the person’s actions as stupid and childish, which we labeled as a bullying tweet. As we can see, the same word can have completely different meanings depending on how it is used, which highlights the difficulty of correctly labeling a cyberbullying dataset. It requires tremendous effort.

TABLE III. SAMPLES OF MANUALLY LABELED TWEETS IN OUR DATASET WHERE THE 0 LABEL INDICATES A NON-BULLYING TWEET WHILE 1 IS FOR BULLYING

Tweets	Label
شهر بدون عيادات.. بمجرد ماتقراها تحس براحة نفسيه	0
شكله عندوا مشاكل نفسيه	1
وجود شخص خبل بحياتك يعتبر نوع من أنواع العلاج النفسي	0
الحمدلله والشكر مستحيل انه انسان عاقل فاهم وفي بخ اللي مسويها خبل ولا بزر	1

### C. Data Pre-processing

The Arabic dataset was preprocessed following standard data mining methods [16], [17], [21], which involved four key steps: cleaning, normalization, stemming, and vectorization.

The cleaning step was crucial to ensure that the dataset contained only relevant and meaningful information for further analysis. We eliminated usernames, as they do not contribute to the sentiment or content of the tweets. Additionally, numbers were removed since they often do not carry significant semantic meaning in the context of text analysis. Null samples and duplicated tweets were also eliminated to ensure data integrity and avoid skewing the analysis. URLs were removed to eliminate any bias or influence that external websites or resources may have on the dataset. Special characters, punctuation marks, and emojis were stripped from the text, as they do not provide valuable information for sentiment analysis and may introduce noise to the data. Finally, English letters were filtered out to focus exclusively on the Arabic text, as the study specifically targeted cyberbullying in the Arabic language.

Following that, we employed normalization techniques to achieve a consistent representation of words, ensuring uniformity in the dataset. We focused on converting different forms of the same word into a common base form. The tweets were normalized and standardized into a unified format. It is worth noting that the dataset consisted of tweets written in both classic Arabic and Modern Standard Arabic (MSA), with variations in dialects based on geographic regions. Furthermore, it was observed that users often substitute diacritics (Tashkeel) with letters, leading to spelling mistakes. For instance, they would write **انو** instead of **انه** and **هاذا** instead of **هذا**. To address this, we applied diacritics and letter normalization techniques to ensure consistency and accuracy in the data. Additionally, we removed stop words, which are commonly used words that carry little semantic meaning. A collection of 750 Arabic stop words compiled by Mohamed Taher Alrefaie was employed for this purpose<sup>1</sup>. Removing these stop words and normalizing the dataset served the dual purpose of reducing dimensionality and avoiding negative impacts on the training process. For a visual reference of the letters used in the samples and their corresponding replacements, please refer to Table IV.

TABLE IV. THE LETTERS USED IN THE SAMPLES AND THEIR REPLACEMENT

Original Letters	Target Letters
ا، آ، إ،	ا
انو	انه
دا، دي، هذي، هذا	هذا
ليه، ليش	لماذا
ة	ه
ى	ا
ؤ، ئ	ء

Finally, we performed vectorization to transform the textual data into numerical representations, enabling the application of machine learning algorithms for classification and analysis. To achieve this, we utilized the CountVectorizer module from the Scikit-Learn library [26]. This powerful tool allowed us to convert each tweet into a matrix of token counts. In simpler terms, CountVectorizer assigns a numerical value to each word in the tweet, indicating the frequency of occurrence. This process effectively creates a numeric representation of the text, which can be easily processed and analyzed by machine learning algorithms. Additionally, we employed the Term-Frequency Times Inverse Document-Frequency (TFIDF) weighting scheme, also provided by Scikit-Learn [26]. TFIDF helps determine the importance and weight of each term within the dataset. This scheme takes into account the frequency of a term within a specific tweet (term frequency) and balances it with the rarity of the term across all tweets (inverse document frequency). As stated by Scikit-Learn, TFIDF can be obtained by:

$$TFIDF(t, d) = TF(t, d) \times IDF(t), \quad (1)$$

$$IDF(t) = \log\left(\frac{n}{df(t)}\right) + 1, \quad (2)$$

where  $n$  is the total number of tweets in the dataset and  $df(t)$  is the dataset frequency of  $t$ . By applying TFIDF, we can normalize the CountVectorizer matrix, providing a more

refined representation of the tweet dataset. These vectorization techniques were essential in transforming the Arabic dataset into a suitable format for machine learning analysis and modeling. By converting the textual data into numerical representations, we enable the algorithms to understand and process the information effectively. This final preprocessing step prepared the dataset for further exploration and utilization of machine learning algorithms to extract valuable insights and classify cyberbullying patterns in the Arabic language.

#### D. Data Classification

In this paper, we utilized Support Vector Machine (SVM) and Multi-layer Perceptron (MLP) [26] as our chosen classification algorithms. SVM is a linear model that constructs a line or hyperplane to separate the data into predefined classes. It aims to find the maximum margin that separates the hyperplane between two data classes, thereby achieving optimal classification performance. One compelling aspect of using SVM in our work is its ability to effectively handle small datasets and provide accurate approximations of the underlying learning patterns. MLP is a type of fully connected feedforward neural network consisting of three layers: the input layer, hidden layer(s), and output layer. For our specific MLP configuration, we employed four hidden layers with 30, 66, 66, and 30 nodes, respectively. Since our dataset only consisted of binary classes, we utilized the logistic activation function.

$$s(x) = \frac{1}{1 + e^{-x}}. \quad (3)$$

One motivating factor for incorporating MLP into our work is its capability to learn complex patterns and relationships in data. Being a fully connected architecture, consisting of multiple layers and a large number of parameters, MLP is a suitable choice for tasks that involve complex data representations with potential non-linear relationships.

#### E. Model Evaluation

In evaluating the performance of a classification model, a range of metrics and techniques are utilized to assess its effectiveness in accurately predicting class labels.

One essential metric is the accuracy metric used to evaluate the performance of a classification model. It measures the overall correctness of the model's predictions by calculating the ratio of correctly classified instances (TP and TN) to the total number of instances (TP, TN, FP, and FN). The accuracy score is computed using the formula:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

Additionally, we assess the model's performance using precision and recall metrics. Precision measures the model's ability to correctly identify true positives among the predicted positive instances. It is calculated by dividing the number of true positives (TP) by the sum of true positives and false positives (FP):

$$Precision = \frac{TP}{TP + FP}. \quad (5)$$

On the other hand, recall, also known as sensitivity or true positive rate, evaluates the model's capability to identify positive instances correctly. It is calculated by dividing the number

<sup>1</sup><https://github.com/mohataher/arabic-stop-words>

of true positives (TP) by the sum of true positives and false negatives (FN):

$$Recall = \frac{TP}{TP + FN}. \quad (6)$$

To provide a balanced assessment of the model's performance, particularly in scenarios with imbalanced class distributions, we employ the F1 score metric. The F1 score combines precision and recall into a single metric, taking into account both the model's ability to correctly identify positive instances and its capability to avoid false positives. It is calculated using the formula:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}. \quad (7)$$

By utilizing these evaluation metrics, including the confusion matrix, accuracy, precision, recall, and F1 score, we can comprehensively evaluate the performance and effectiveness of our classification model in accurately predicting class labels within the given dataset.

#### IV. RESULTS AND DISCUSSIONS

In this section, we present a comprehensive evaluation of our proposed approach. We begin by outlining the research questions that we aim to answer, followed by the experimental setup, and we conclude with the results and the findings.

TABLE V. MODEL PERFORMANCE

Split#	ACC	Precision	Recall	F1
MLP				
1	89	87	92	89
2	88	88	89	89
3	89	88	90	89
4	89	88	90	89
5	91	90	92	91
AVG	89	88	90	89
SVM				
1	91	87	94	91
2	93	92	94	93
3	91.7	90	95	92
4	91.8	91	93	92
5	92	89.7	94.6	92
AVG	92	90	94	92

##### RQ1. How accurately can we classify cyberbullying?

*Experimental Setup* Our machine-learning code was implemented using Python 3.7, and we utilized the Scikit-Learn library [26] for building the classification model. The experiments were conducted on a Dell Inspiron 5406 laptop equipped with a 2.8 GHz 4-Core Intel Core i7 processor and 16 GB of memory.

To ensure fairness in training the model on different samples, specifically the classes of 0's and 1's, we examined the entropy of the datasets before initiating the training process. The uniformity, as a measure of data entropy, was evaluated based on the Hamming weights of the dataset's responses. The uniformity score ( $U_s$ ) was calculated using the following formula:

$$U_s = \frac{1}{C} \sum_{i=1}^C r_i \times 100, \quad (8)$$

where  $r_i$  represents the class bit generated when the input dimensions are from the  $s$ -th tweet set or the  $s$ -th sample, and

$C$  denotes the total number of tweets in a file. By examining the uniformity scores, we ensured that both classes had a balanced representation within the training data, minimizing the potential bias towards any particular class. This step was crucial to maintain fairness and prevent the model from being biased towards the majority class during the training phase.

To address our research question, we divided the dataset into an 80% training set and a 20% testing set. We applied two classification models, namely Support Vector Machine (SVM) and Multi-Layer Perceptron (MLP) [26]. In the MLP classifier, we designed the architecture with 192 neurons in the input layer and 4 hidden layers. To optimize the model's performance, we utilized the Adam optimizer and employed the logistic activation function. These choices were made to enhance the model's ability to capture complex patterns and relationships in the data during the training process

*Results and Findings* Table V displays the model performance on five different 80/20 splits of the dataset, as well as the average performance on classifying the testing dataset, which constitutes 20% of the entire dataset. The table provides a comprehensive overview of the performance metrics for each split, allowing for a comparison of the models' consistency across different subsets of the data. The inclusion of multiple splits helps to mitigate the potential impact of dataset variability and provides a more robust evaluation of the models' performance. By averaging the results across these splits, we obtain a more reliable estimation of the models' general performance on unseen data. Overall, the results from Table V demonstrate that SVM outperformed MLP in the classification task, achieving better accuracy, F1 score, precision, and recall. These findings indicate that SVM was more effective in accurately predicting class labels in the testing dataset, making it a favorable choice for the classification task at hand.

##### RQ2. How does ArCyb compare to the state of the art?

*Experimental Setup* Arabic language sentiment analysis is a challenging task that requires significant effort to achieve high prediction rates due to the complexity of the language and the need for a well-labeled dataset. To provide a comprehensive evaluation of our model, we plan to compare it against state-of-the-art models developed. Specifically, we will evaluate our model against the models proposed by Almutiry and Fattah, Almutairi and Alhagry [19], [20], who achieved accuracies of 85% and 82%, respectively, in their cyberbullying models. To ensure a fair comparison, we will replicate their approach, including their text preprocessing and model architecture, to evaluate the effectiveness of our model in detecting cyberbullying in Arabic language texts.

*Results and Findings* Upon analyzing the work of Almutiry and Fattah [19] and Almutairi and Alhagry [20], we found that both studies have invested considerable effort in building their models and implementing preprocessing methodologies. A comparison of their approaches is presented in Table VI. We can observe that our approach outperformed all other approaches. We believe our approach performed better due to several factors, but one potential key difference lies in the stemming step during the preprocessing stage. Specifically, Almutairi and Alhagry [20] did not apply stemming to their data, whereas Almutiry and Fattah [19] employed light stemmer

and Khoja stemmer. In our approach, we utilized AraBERT [27] for stemming. The Khoja stemmer and the light stemmer are rule-based stemmers that rely on predetermined rules to remove inflectional endings from Arabic words, resulting in the base form of the word. The effectiveness and precision of these stemmers depend on the thoroughness of the rules and the complexity of the Arabic inflectional system. In contrast, AraBERT is a machine learning model trained on a large dataset of Arabic text. This enables AraBERT to perform automated and adaptable stemming by considering the context and relationships between words in both left-to-right and right-to-left directions. By understanding the surrounding words, AraBERT gains a better understanding of the meaning of the text. Through our evaluation, we aim to investigate the impact of different stemming approaches on the performance of the models.

TABLE VI. COMPARING ARCYB WITH ALMUTAIRI AND ALHAGRY [20], ALMUTIRY AND FATTAH [19]

Author	ACC	Precision	Recall	F1-score
Almutiry and Fattah [19]	90	88	92	90
Almutairi and Alhagry [20]	89	86	92	89
Our approach	92	90	94	92

TABLE VII. VALIDATING ARCYB MODEL USING AJGT DATASET

Model	Accuracy	Precision	Recall	F1
ArCyb	91	90	92	91
[28]	88.72	92	84	88.27

### RQ3. Can our ArCyb Machine-Learning approach be used on similar problems?

*Experimental Setup* For this research question, we aim to evaluate whether our approach can be applied to similar problems, such as sentiment analysis. To achieve this, we will compare the performance of our approach with an established Arabic sentiment analysis model. We will utilize the Arabic Jordanian General Tweets (AJGT) dataset obtained from Alomari et al. [28]. The AJGT dataset consists of 1800 samples, each labeled with either a positive or negative sentiment. By using the AJGT dataset, our goal is to assess and compare the predictive capabilities of our model against the state-of-the-art Arabic sentiment analysis model. This comparative analysis will allow us to evaluate the performance, accuracy, and reliability of our proposed approach. Additionally, it serves as a benchmark for determining the effectiveness of our model in capturing and understanding the sentiments expressed in Arabic language text. We will apply the same preprocessing methods to the AJGT dataset as described in III. Splitting the dataset into 80% for training and 20% for testing.

*Results and Findings* The performance evaluation results presented in Table VII demonstrate that our MLP model surpassed the performance of the original model proposed in [28], achieving an accuracy of 91% compared to the original model's accuracy of 88%. This outcome suggests that our approach has the potential to outperform existing models in various Arabic classification problems, extending beyond the domain of cyberbullying. By demonstrating superior performance in this comparative analysis, our model showcases its effectiveness in accurately classifying Arabic text across different contexts and applications. These findings highlight

the versatility and generalizability of our approach, making it a promising solution for a wide range of classification tasks in the Arabic language.

### RQ4. What insights can ArCyb tell us about Cyberbullying on Twitter?

*Experimental Setup* To validate the effectiveness and applicability of our model in classifying unlabelled data, we will utilize the same set of 16 keywords that were used in the original model. We will collect raw unlabelled data from the period of 2013 to 2022, consisting of 1000 samples for each keyword. This extensive dataset will enable us to analyze and quantify the prevalence and occurrences of bullying events over the past ten years. By applying our model to this unlabelled data, we aim to gain valuable insights into the bullying rate and trends, providing a deeper understanding of the dynamics and impact of bullying during the studied period.

*Results and Findings* Fig. 3 display the bullying rates in the last decade, which show an obvious increase in the bullying rate by 35.9% between the years 2013-2022.

We have further investigated the data to identify the most frequent words that occurs in the bullying samples. These words are not necessary bullying words but they were used in the same tweet that is classified as bullying based on it's context. The most frequent words are displayed in Fig. 2. Here are a few noteworthy examples from our findings:

In 2013, tweets related to Alittihad FC revealed dissatisfaction among fans regarding the team's performance and the management under the leadership of Mohammad Alfayez. Bullying tweets targeting the team's performance, players, and management decisions prominently featured the name "Mohammad Alfayez".

In 2014, there was a significant social media backlash against the prank show "Ramez the Sea Shark" hosted by Ramez Galal. Many viewers found the show unfunny and insulting to the guests, leading to the creation of memes that ridiculed the show. The show's name, "Ramez" and "sea", were frequently mentioned in bullying tweets.

The emergence of the Houthi movement in 2015 sparked a surge in hateful tweets and cyberbullying directed towards the group. Social media users expressed offensive and derogatory opinions, leading to an ongoing trend of bullying against the Houthi movement throughout the years, including 2022.

In 2019, Shawarmer, a popular fast food chain, posted a tweet that was deemed disrespectful to Alhilar FC, a prominent football club. This incident resulted in a hashtag campaign bullying Shawarmer's products as a form of retaliation.

In 2022, the Africa Cup of Nations (AFCON) generated significant attention on social media, particularly matches involving Algeria, Cameroon, Egypt, and Senegal. During the final match between Egypt and Senegal, an incident occurred where Senegalese fans pointed lasers at Egyptian player Mohamed Salah during penalty kicks. This incident sparked outrage on the internet and became a trending topic, leading to an influx of bullying-related hashtags.

Throughout the years 2013-2022, the top Saudi Arabian football clubs including Ittihad, Alhilar, Alahli, and Alnasser,



Fig. 2. Words mentioned in bullying samples.

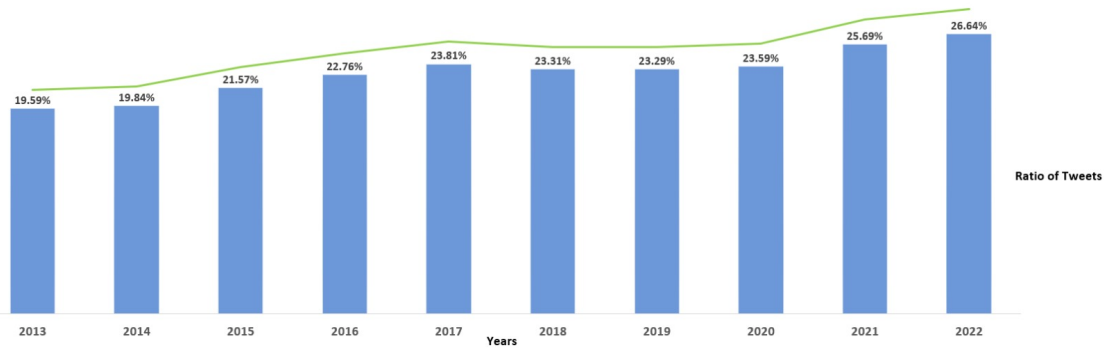


Fig. 3. Cyberbullying rate in the last decade.

were consistently mentioned in the bullying samples. Additionally, the names of famous football players were frequently targeted, highlighting football as a hot topic for cyberbullying.

These examples highlight the diversity of bullying topics and events observed in the collected tweets from 2013 to 2022, providing valuable insights into the dynamics of online bullying and its association with various social, cultural, and sporting phenomena.

## V. CONCLUSION

In this research, we undertake the task of building our dataset from scratch. We start by collecting raw data, obtaining a total of 4,140 samples. To ensure a focused collection, we specify 16 bullying terminologies and use them as keywords to pull relevant data from Twitter via the Twitter API. Subsequently, we form a group consisting of three cybersecurity specialists who manually label the samples to ensure accurate annotation. After the dataset collection and labeling process, we proceed to the preprocessing phase. This involves several steps, including data cleaning, normalization, stemming, and vectorization. These steps are necessary to prepare the data for classification. Using both MLP and SVM classifiers, we conduct classification experiments on the preprocessed dataset.

The results demonstrate an accuracy of 89% for MLP and 92% for SVM. These promising performance metrics validate the effectiveness of our approach in classifying cyberbullying instances. Additionally, we seek to assess the accuracy and predictive capabilities of our model by gathering a large dataset consisting of 160,000 raw tweets spanning the years 2013 to 2022. Through analysis, we identify the most frequent words associated with bullying, which reflect specific events that occur during different periods of time. Notably, our findings indicate a significant increase in the bullying rate, with an annual growth rate of 35.9%. These findings highlight the effectiveness and relevance of our model in addressing the challenges of cyberbullying detection and classification. Furthermore, our analysis of the collected tweets provides valuable insights into the evolving landscape of online bullying, indicating the need for continued efforts to combat this issue.

## ACKNOWLEDGMENTS

This work was funded by the University of Jeddah, Jeddah, Saudi Arabia, under grant No. (UJ-21-IMT-10). The authors, therefore, acknowledge with thanks the University of Jeddah technical and financial support. The authors would like to thank



Suhail A. Segnawi, and Aws H. Aljahdli for their help in maintaining the dataset.

#### REFERENCES

- [1] R. S. Tokunaga, "Following you home from school: A critical review and synthesis of research on cyberbullying victimization," *Computers in human behavior*, vol. 26, no. 3, pp. 277–287, 2010.
- [2] H. Dani, J. Li, and H. Liu, "Sentiment informed cyberbullying detection in social media," in *Joint European conference on machine learning and knowledge discovery in databases*. Springer, 2017, pp. 52–67.
- [3] A. M. Al-Zahrani, "Cyberbullying among saudi's higher-education students: Implications for educators and policymakers." *World Journal of Education*, vol. 5, no. 3, pp. 15–26, 2015.
- [4] J. J. Dooley, J. Pyżalski, and D. Cross, "Cyberbullying versus face-to-face bullying: A theoretical and conceptual review," *Zeitschrift für Psychologie/Journal of Psychology*, vol. 217, no. 4, pp. 182–188, 2009.
- [5] C. Langos, "Cyberbullying: The challenge to define," *Cyberpsychology, behavior, and social networking*, vol. 15, no. 6, pp. 285–289, 2012.
- [6] J. W. Patchin and S. Hinduja, "Measuring cyberbullying: Implications for research," *Aggression and Violent Behavior*, vol. 23, pp. 69–74, 2015.
- [7] Economic and S. R. Council, "'how to use social media'," Oct. 14, 2021, accessed Mar. 2, 2022. [Online]. Available: <https://www.ukri.org/councils/esrc/impact-toolkit-for-economic-and-social-sciences/how-to-use-social-media/choosing-what-social-media-you-use/>
- [8] A. Alasem, "egovernment on twitter: The use of twitter by the saudi authorities," *Electronic Journal of e-Government*, vol. 13, no. 1, pp. pp67–73, 2015.
- [9] B. Saberi and S. Saad, "Sentiment analysis or opinion mining: a review," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 7, no. 5, pp. 1660–1666, 2017.
- [10] G. Julian, "What are the most spoken languages in the world," *Retrieved May*, vol. 31, p. 2020, 2020.
- [11] M. Abdul-Mageed, M. Diab, and M. Korayem, "Subjectivity and sentiment analysis of modern standard arabic," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 2011, pp. 587–591.
- [12] J. Wiebe, R. Bruce, and T. P. O'Hara, "Development and use of a gold-standard data set for subjectivity classifications," in *Proceedings of the 37th annual meeting of the Association for Computational Linguistics*, 1999, pp. 246–253.
- [13] R. M. Duwairi and I. Qarqaz, "Arabic sentiment analysis using supervised classification," in *2014 International Conference on Future Internet of Things and Cloud*. IEEE, 2014, pp. 579–583.
- [14] A. Shoukry and A. Rafea, "Sentence-level arabic sentiment analysis," in *2012 International Conference on Collaboration Technologies and Systems (CTS)*. IEEE, 2012, pp. 546–550.
- [15] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? sentiment classification using machine learning techniques," *arXiv preprint cs/0205070*, 2002.
- [16] M. N. Al-Kabi, A. H. Gigieh, I. M. Alsmadi, H. A. Wahsheh, and M. M. Haidar, "Opinion mining and analysis for arabic language," *IJACSA International Journal of Advanced Computer Science and Applications*, vol. 5, no. 5, pp. 181–195, 2014.
- [17] H. Al-Rubaiee, R. Qiu, and D. Li, "Identifying mubasher software products through sentiment analysis of arabic tweets," in *2016 International Conference on Industrial Informatics and Computer Systems (CIICS)*. IEEE, 2016, pp. 1–6.
- [18] B. Y. AlHarbi, M. S. AlHarbi, N. J. AlZahrani, M. M. Alsheail, J. F. Alshobaili, and D. M. Ibrahim, "Automatic cyber bullying detection in arabic social media," *Int. J. Eng. Res. Technol.*, vol. 12, no. 12, pp. 2330–2335, 2019.
- [19] S. Almutiry and M. Abdel Fattah, "Arabic cyberbullying detection using arabic sentiment analysis," *The Egyptian Journal of Language Engineering*, vol. 8, no. 1, pp. 39–50, 2021.
- [20] A. R. Almutairi and M. A. Al-Hagery, "Cyberbullying detection by sentiment analysis of tweets' contents written in arabic in saudi arabia society," *International Journal of Computer Science & Network Security*, vol. 21, no. 3, pp. 112–119, 2021.
- [21] A. Al Sallab, H. Hajj, G. Badaro, R. Baly, W. El-Hajj, and K. Shaban, "Deep learning models for sentiment analysis in arabic," in *Proceedings of the second workshop on Arabic natural language processing*, 2015, pp. 9–17.
- [22] A. Dahou, S. Xiong, J. Zhou, M. H. Haddoud, and P. Duan, "Word embeddings and convolutional neural network for arabic sentiment classification," in *Proceedings of coling 2016, the 26th international conference on computational linguistics: Technical papers*, 2016, pp. 2418–2427.
- [23] S. Tartir and I. Abdul-Nabi, "Semantic sentiment analysis in arabic social media," *Journal of King Saud University-Computer and Information Sciences*, vol. 29, no. 2, pp. 229–233, 2017.
- [24] S. R. El-Beltagy, T. Khalil, A. Halaby, and M. Hammad, "Combining lexical features and a supervised learning approach for arabic sentiment analysis," in *International Conference on Intelligent Text Processing and Computational Linguistics*. Springer, 2016, pp. 307–319.
- [25] E. Othman, K. Shaalan, and A. Rafea, "Towards resolving ambiguity in understanding arabic sentence," in *International Conference on Arabic Language Resources and Tools, NEMLAR*. Citeseer, 2004, pp. 118–122.
- [26] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [27] W. Antoun, F. Baly, and H. Hajj, "AraBERT: Transformer-based model for Arabic language understanding," in *Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection*. Marseille, France: European Language Resource Association, May 2020, pp. 9–15. [Online]. Available: <https://aclanthology.org/2020.osact-1.2>
- [28] K. M. Alomari, H. M. ElSherif, and K. Shaalan, "Arabic tweets sentimental analysis using machine learning," in *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*. Springer, 2017, pp. 602–610.

# Development of a Touchless Control System for a Clinical Robot with Multimodal User Interface

Julio Alegre Luna, Anthony Vasquez Rivera, Alejandra Loayza Mendoza,  
Jesús Talavera S., Andres Montoya A.  
Universidad Nacional de San Agustín de Arequipa, Perú

**Abstract**—This article introduces the development of a multimodal user interface for touchless control of a clinical robot. This system seamlessly integrates distinct control modalities: voice commands, an accelerometer-embedded gauntlet, and a virtual reality (VR) headset to display real-time robot video and system alerts. By synergizing these control approaches, a more versatile and intuitive means of commanding the robot has been established. This assertion finds support through comprehensive assessments conducted with both seasoned professionals and novices in the domain of clinical robotics, all within a controlled experimental setting. The diverse array of test results unequivocally demonstrate the system's efficacy. They substantiate the system's ability to proficiently govern a robotic arm in the clinical environment. The user interface's usability is measured at an impressive 90.2 on the system usability scale, affirming its suitability for robotic control. Notably, the interface not only offers comfort but also intuitiveness for operators of varying levels of expertise.

**Keywords**—Multimodal user interface; human-robot interaction; clinical robot

## I. INTRODUCTION

Robotic systems are progressively assuming greater significance in the development of conventional human activities, especially within the realm of healthcare, owing to their array of merits. These encompass heightened efficiency and precision, risk mitigation, and enhanced patient comfort [1]. These systems find versatile utility across a spectrum of clinical applications, spanning from surgical procedures and rehabilitation to technologies tailored for the aid of the elderly or disabled [2], [1]. Yet, the management of these systems within a clinical milieu poses intricate challenges. Conventional interfaces like buttons and joysticks, while conventionally employed, engender potential infection hazards and present difficulties for patients possessing restricted mobility or dexterity [3].

In response to these challenges, researchers have embarked upon the exploration of novel paradigms for robotic control that are imbued with heightened intuition, naturalness, and touchlessness [4], [5]. An auspicious avenue in this endeavor is the adoption of multimodal user interfaces, which amalgamate an assortment of control methodologies, thus engendering a more adaptable system [6]. Multimodal user interfaces empower users to seamlessly transition between diverse control modes contingent upon their proclivities or the specific task at hand [6]. As an illustration, a user might seamlessly oscillate between voice commands to oversee the robot's locomotion, while seamlessly transitioning to gesture-based control for tasks demanding precision in manipulation [7].

This article introduces a multimodal user interface developed for touchless control of a clinical robot. This system seamlessly integrates two distinct control methodologies: voice commands, and an accelerometer-embedded gauntlet. The voice command system empowers users to steer the robot through spoken directives, while the accelerometer-equipped gauntlet detects user-initiated gestures. Finally, the virtual reality headset allows the operator to visualize in real time the video captured by the webcam, and also allows the system to display on screen different alerts triggered by the two methods mentioned above. This system boasts a spectrum of prospective applications within varied clinical contexts. For instance, it could find utility in surgical settings, enabling surgeons to orchestrate robot movement while maintaining a sterile environment. Similarly, within rehabilitation realms, patients might exercise dominion over robotic devices using their voice or gestures. Furthermore, the system's utility is magnified for individuals with restricted mobility or dexterity, as it allows them to exert control over the robot devoid of physical interaction. To the best of current knowledge, this represents the maiden multimodal user interface tailored for touchless control of a clinical robot, concomitantly amalgamating voice commands, accelerometer-equipped gauntlet, and a VR headset that displays the developed user interface. The system's architectural blueprint prioritizes user-friendliness, safety, and reliability, buttressed by a series of meticulously devised experiments aimed at scrutinizing its efficacy in robot manipulation.

The subsequent sections of this article are structured as follows: In Section II, Related Works, prior research concerning voice-controlled systems, gesture recognition systems, and multimodal user interfaces for robot and robotic arm control is comprehensively surveyed. Section III, Experimental Development, an exhaustive account is provided regarding all employed electronic components, the clinical robotic platform, the implementation of the voice control system, the accelerometer-embedded gauntlet, and the associated software architecture conceived for this integrated system. The subsequent segment, Section IV, Results and Discussions, unveils the empirical outcomes derived from the conducted experiments, accompanied by the ensuing discussion arising from their interpretation. Ultimately, in Section V, this document culminates as conclusions are drawn and prospective avenues of research are deliberated upon.

## II. RELATED WORKS

### A. Voice-Control Systems

Sagar's article [8] reviewed the current status of speech recognition systems. In addition, the potential industrial applications of speech recognition technology, such as public safety solutions, were discussed. Furthermore, the article delves into the future scope of voice recognition, with the potential for artificial intelligence to reshape how we interact with devices [9]. In another study [10], a voice recognition control system for a robot is delineated, designed to operate effectively in noisy environments. This system employs generalized sidelobe canceller techniques resilient to outliers, noise suppression in the feature space, and reverberation mitigation. The article also delves into obstacle detection, local map design, as well as target search and avoidance behaviors using fuzzy decision-making. The system's efficacy is evaluated on a communication robot deployed within a real noisy environment. The article also contemplates the integration of robust voice recognition and navigation systems for autonomous navigation within unfamiliar surroundings.

In another study [11], a system is proposed that provides a mobile robot with the ability to separate simultaneous sound sources; an array of microphones is used along with a dedicated real-time implementation of geometric source separation and a post-filter that provides us with further reduction of interference from other sources. The work of [12] discusses the creation of target-seeking and avoidance behaviors employing fuzzy decision-making. The author in [13] introduces a method for selecting an appropriate behavior from numerous primitive behaviors using a fuzzy decision-maker. author in [14] describes an obstacle detection method and local map design utilizing an array of ultrasonic sensors. Finally, [15] introduces a novel approach to voice recognition in noisy environments, grounded in multi-condition training techniques, maximum likelihood linear regression, and missing feature theory.

### B. Gesture Recognition Systems

The article [16] introduces a human-computer interaction (HCI) model based on somatosensory interaction for robotic arm manipulation. The model utilizes a 3D SSD architecture for gesture and arm movement localization and identification, coupled with the Dynamic Time Warping (DTW) template matching algorithm for dynamic gesture recognition. Interactive scenarios and modes are designed for experimentation and implementation, with virtual experimental results demonstrating the method's efficacy. In [17], a real-time hand gesture recognition system is presented for controlling mobile robots using vision sensors. The system employs image processing techniques to extract the center of mass and features of a red glove worn by the user. These features are then used to control the robot's movements. The design of the mobile robot is uncomplicated and tailored for the system, consisting of three layers with a 4 cm separation to accommodate circuit placement. The system employs a motor control circuit and a PIC18F452 microcontroller control circuit. Additionally, the system incorporates XBee wireless transmitter and receiver modules for data transmission. The system employs color filtering to extract the red glove's shape and spot size filtering to eliminate objects below a certain size [18].

The article [19] presents a gesture recognition system for interacting with computers in dynamic environments. The system employs image processing techniques for hand gesture detection, segmentation, tracking, and recognition, transforming them into meaningful commands. The proposed interface finds applicability across diverse domains like image navigation and gaming. Real-world scenario testing exhibited effective performance in low-noise environments and balanced lighting conditions. The designed gesture vocabulary can be expanded to control different applications, enhancing adaptability in human-computer interaction. This work is aligned with research in the field of human-computer interaction and gesture recognition. The article [20] pertains to the realm of human-robot interaction, focusing on real-time hand gesture recognition to enhance human-robot interaction within dynamic environments. Enhanced classifiers are employed for hand detection and static gesture recognition, while a Bayesian classifier is utilized for dynamic gesture recognition. Additionally, the system incorporates contextual information, such as human face detection and tracking, to enhance robustness and speed. Relevant works utilizing contextual information to improve the accuracy and speed of gesture recognition systems are also referenced. The proposed system's validation is conducted on actual video sequences, achieving a recognition rate of 70% for static gestures and 75% for dynamic gestures, operating at varying speeds of 5-10 frames per second.

### C. Multimodal User Interfaces

The document [21] presents work related to human-robot collaboration (HRC) in manufacturing, specifically in assembly tasks. The challenges and limitations of existing HRC systems are discussed, including issues such as lack of adaptability and flexibility in task programming, along with the need to ensure human safety in the working environment. The article proposes a solution based on the utilization of function blocks and intuitive multimodal control to enhance flexibility and adaptability in complex assembly tasks. Concepts of multimodal control, function blocks, and their application in human-robot collaboration within manufacturing are thoroughly examined. Conversely, the article [22] delves into a usability study of three interfaces designed to present search engine results on the Internet. The study compared a text-only interface with two others that combined text, visual metaphors, and voice messages. Results indicated that the multimodal interfaces were more usable than the text-only interface. In a third work, Lunghi's article [23] details the design and software engineering process behind the development of a multimodal Human-Robot Interface (HRI) for intervention with a cooperative team of robots. The operator gains the capability to enter the control loop between the HRI and the robot, customizing control commands in accordance with the operation.

### D. Robots in the Clinical Environment

Poirier's paper [24] presents the design and preliminary evaluation of a voice command system prototype for the control of assistive robotic arms' movements; the prototype of the voice command interface developed is first presented, followed by two experiments with five able-bodied subjects in order to assess the system's performance and guide future development. In the work of Morgan et al. [25], a comprehensive literature review is presented, focusing on the utilization

of robots within the realm of healthcare. The study identifies ten primary roles that robots can undertake in clinical settings, encompassing surgery, rehabilitation and mobility, radiotherapy, social assistance, telepresence, pharmacy, disinfection, delivery and transportation, image intervention, and assistance. Furthermore, the article underscores robots' potential to adapt to the dynamic demands of healthcare, including those that arise during pandemics.

In Peter's study [26], a novel multimodal human-machine interface system is developed using combinations of electrooculography (EOG), electroencephalography (EEG), and electromyogram (EMG) to generate numerous control instructions; the results indicate that the number of system control instructions is significantly greater than achievable with any individual mode. In other paper [27], an interface centered on the deployment of the Leap Motion (LM) controller is examined. This interface facilitates the real-time tracking of a surgeon's hand position and orientation, capturing nuanced finger gestures and movements, which are subsequently relayed to a computer. Subsequently, a surgical robotic arm is manipulated using data gleaned from the LM controller, data that is systematically classified through programming. Beyond the capabilities attributed to the LM controller, attributes like its cost-effectiveness, acceptable precision, and high-speed data processing have rendered it a feasible and efficient tool for application.

### III. METHODOLOGY

The proposed system enables the user to control a contactless robotic arm through a multimodal user interface. Two control methods are integrated: voice commands, and hand gestures. The voice command system empowers users to steer the robot through spoken directives, while the accelerometer-embedded gauntlet detects user-initiated gestures. A VR headset that displays the developed user interface. Fig. 1 illustrates the block diagram of the proposed system, which is subsequently elaborated upon in each stage.

#### A. Hardware Components

The hardware components utilized in this study encompass a Raspberry Pi, an Arduino Nano with WiFi module, a microphone, an accelerometer-embedded gauntlet, and a webcam. The Raspberry Pi 4 serves as the central processing unit of the system. The Arduino is employed to manage the motor drivers of the robotic arm. The microphone, along with its associated circuitry, is employed for voice command recognition and transmission to the Raspberry Pi. The accelerometer-embedded gauntlet captures hand gestures executed by the user. Lastly, the webcam connected to the Raspberry Pi detects the user's facial features.

1) *Microphone Circuit:* The selected transducer type is a microphone, which is connected to an amplification stage (MAX4455 amplifier) to condition the signal to the desired voltage level, ranging between 0 and 5V. The microphone captures sound waves and converts them into an electrical signal, which is then transmitted to the Raspberry Pi microprocessor. Positioned between the amplification stage and the microprocessor is an analog-to-digital conversion stage (ADS1115 converter). This conversion stage is crucial as it

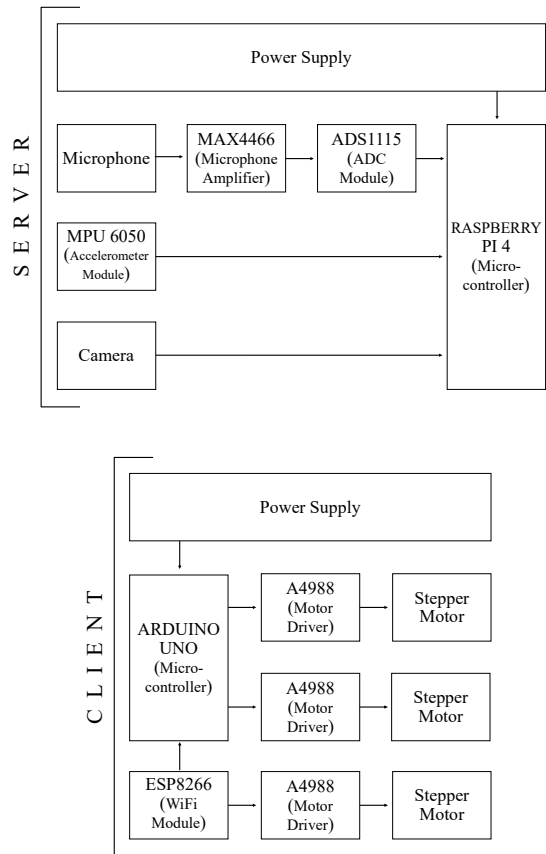


Fig. 1. Block diagram of the complete system.

enables the collection of analog signals and their subsequent processing in a digital-origin microprocessor.

In Fig. 2, the two pins of the microphone are connected, one to the amplification stage and the other to GND. The amplifier is configured in a non-inverting setup. The amplifier's output is connected to pin 1 of the ADC. Pin 7 of the ADC is connected to GPIO9 on the Raspberry Pi, serving as the data transmission pin. Lastly, pin 8 is connected to GPIO10 on the Raspberry Pi, serving as the clock signal pin.

2) *Glove Circuit with Accelerometer:* Accelerometers are devices that measure acceleration force in units of gravity (g) and can measure in one, two, or three planes (X, Y, and Z). The chosen module for this stage is the MPU-6050, which integrates a MEMS accelerometer and a MEMS gyroscope on a single chip. This module is installed in a glove worn by the operator of the robotic arm, capturing hand movements as well as any rotations they perform.

In Fig. 3, the GND pin of the MPU-6050 module is connected to the circuit's ground, while the VDD pin is linked to the voltage output of the Raspberry Pi 4. The SDA pin transmits accelerometer module data to the Raspberry Pi and is connected to GPIO3. The SCL pin of the MPU-6050 module transfers the clock signal to the module and is linked to GPIO5.

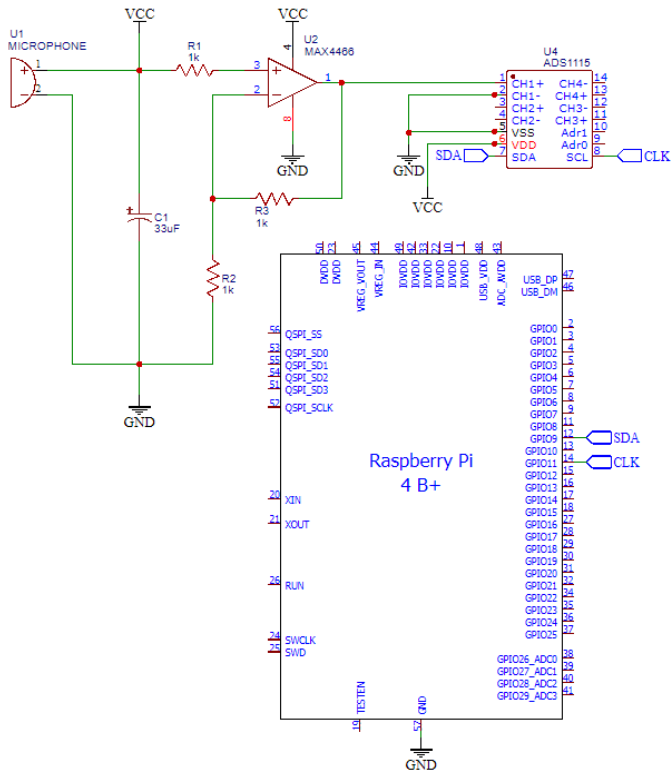


Fig. 2. Connection circuit between the microphone and the Raspberry Pi microprocessor.

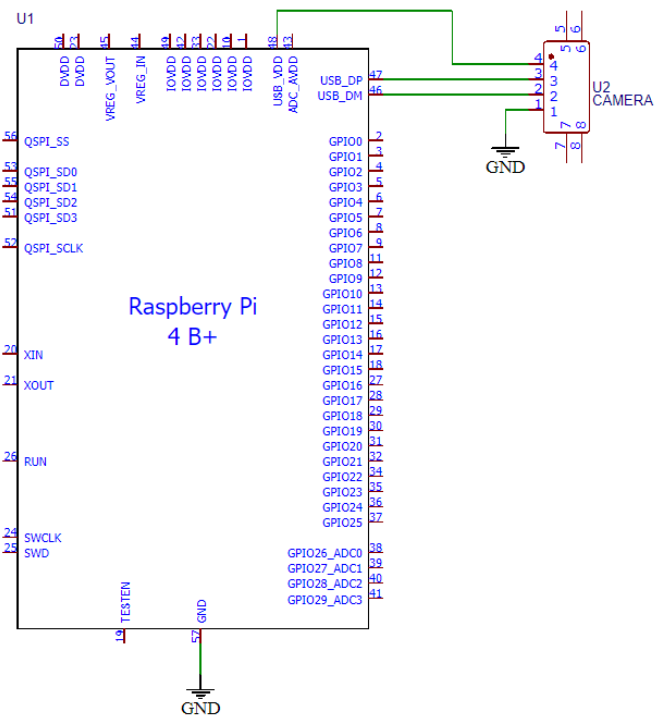


Fig. 4. Connection circuit between the camera and the Raspberry Pi microprocessor.

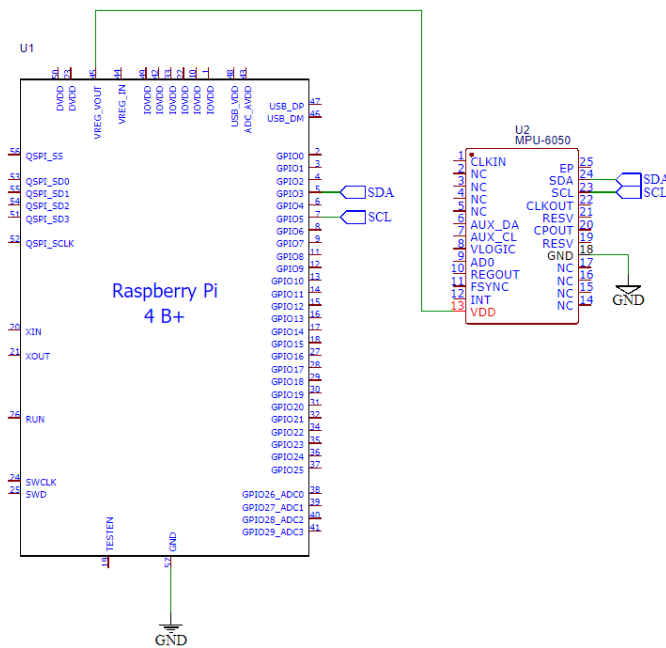


Fig. 3. Connection circuit between the accelerometer of the glove and the Raspberry Pi microprocessor.

3) *Webcam Circuit:* The operation of this stage is straightforward. A webcam is used to transmit video of the robot, which will serve as feedback to the multimodal system. In Fig. 4, the Logitech C922 camera is connected to the Raspberry Pi 4 via its USB port.

4) *Wireless Communication:* In this stage, there are two components. On one side, there's the Raspberry Pi 4, which comes equipped with integrated Wi-Fi. This Wi-Fi functionality is utilized to create a server, enabling the robotic arm to connect to it as a client. As for the robotic arm segment, an Arduino Uno is employed for control. However, since the Arduino Uno doesn't have a built-in Wi-Fi module, an external Wi-Fi module, specifically the ESP8266, is utilized for this purpose.

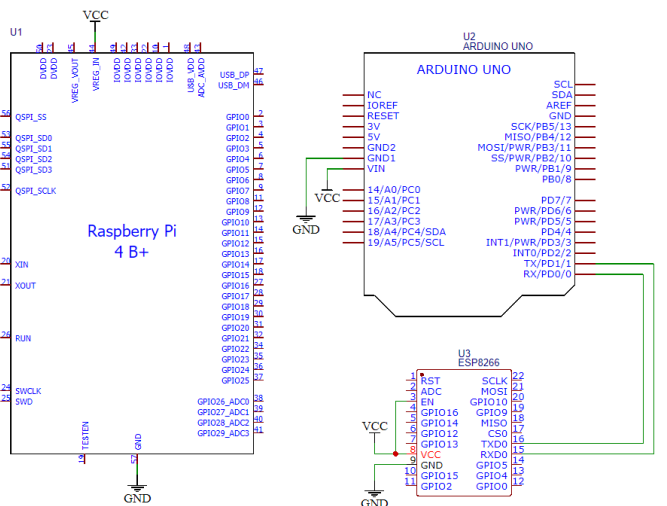


Fig. 5. Wireless communication connection circuit.

In Fig. 5, the Raspberry Pi 4, Arduino, and the ESP8266 module share common VCC and GND connections. The

Arduino and ESP8266 are linked through the TX and RX transmission pins.

5) *Connection of the Robotic arm:* The testing robotic arm has 3 degrees of freedom (3-DOF), which is why 3 stepper motors and 3 drivers are employed to control its movements. These components are connected to the Arduino to issue commands for their respective functions. The A4988 drivers are chosen due to their high reliability in tasks of this nature.

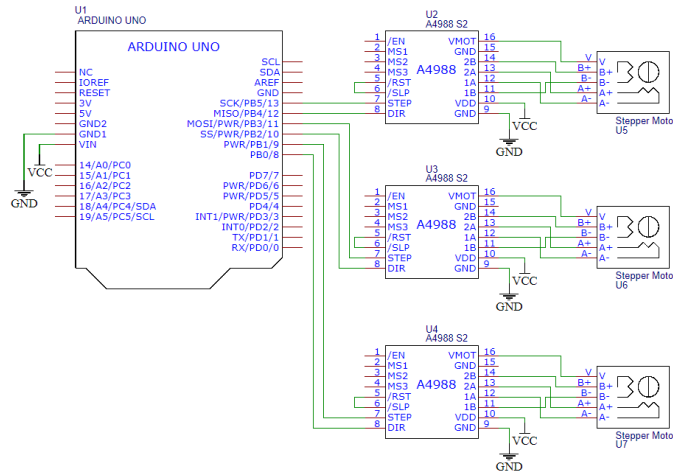


Fig. 6. Robotic arm connection circuit.

In Fig. 6, the Arduino and the motor drivers share the VCC and GND power supply. From the motor drivers, two pins are used to connect to the Arduino: the STEP and DIR pins. These pins determine the number of steps and the direction of rotation, respectively. The pins of the first driver are connected to Arduino pins 12 and 13, the pins of the next driver are connected to pins 10 and 11, and finally, the pins of the third driver are connected to pins 8 and 9 of the Arduino. Each driver is linked to a 4-wire stepper motor, with the 4 wires connected to A+, A-, B+, and B- pins.

**B. Software Components**

The system was developed using the Python programming language due to its extensive library support and versatility for programming innovative systems. Python was utilized to integrate the various software components of the system and to control the robot based on user inputs. Additionally, several software components were employed to make the system function effectively. On the other hand, the Arduino was programmed using its own platform and libraries for motor drivers. The following are the most significant details for this purpose.

1) *Voice Interaction:* For voice recognition, Google Speech-to-Text was employed to transcribe the voice commands issued by the user. It was implemented using the Google Cloud API and integrated into the Python code executed on the Raspberry Pi 4. Voice command language was preferred as the input to the system because it allows a more intuitive interaction to the human’s natural being [28]. The commands used and recognized by the system are displayed in Table I. An indicator provides visual information to the operator of the action being executed. During active navigation mode, the

indicator is illuminated according to the Table I. This lets the operator know which command is currently being executed, as well as whether the spoken command was successfully acknowledged.

TABLE I. DESCRIPTION OF INTEGRATED VOICE CONTROL COMMANDS

Voice Command	Indicator	Description
Start	● ●	Activates glove gesture detection
Translation	○ ●	Initiates translational navigation mode
Rotation	● ○	Initiates rotational navigation mode
Move to	● ●	Opens the options assignment display
Cancel	● ○	Closes the options assignment display
Stop	● ●	Deactivates the selected navigation mode
End	● ●	Concludes the current interaction task

2) *Glove Gesture Interaction:* Control through the glove is achieved by integrating an accelerometer which captures the degrees of inclination of the hand in its different axes (X, Y and Z). The multimodal system allows the operator to execute different commands to the robotic arm thanks to the integration of voice commands, some examples are the function of pausing the sending of accelerometer data to the Raspberry Pi, this allows the operator to rest momentarily or move the hand without worrying that the robot will recreate this movement; Other examples are the rotation and translation commands that allow the robotic arm to move according to the indication executed and the hand movement performed.

3) *User Interface Display:* The user interface collects the webcam video sent by the Raspberry Pi 4 from the server, this video is processed and the spoken command indicator is added so that the operator can realize that it worked correctly; to process the video, the OpenCv library belonging to the Python programming language was used. OpenCV is a powerful computer vision library that was used to detect hand gestures performed by the user. In Fig. 7, you can see an image of the processed video.



Fig. 7. User interface develops visualized in the virtual reality headset.

**C. Testing**

To assess the performance of the proposed multimodal system, a series of experiments were conducted involving 12 participants. Each participant was assigned a set of tasks to perform with the robot, including moving the robotic arm to a

desired position, touching specific elements in the environment with the end effector, among others. Fig. 8 depicts the testing scenario used to evaluate the system's effectiveness.



Fig. 8. Testing environment for the robotic arm.

Participants were instructed to use each of the three control methods (voice commands, glove gestures, and computer vision) individually and in combination to control the robot. The sequence of method usage was randomized to mitigate order effects. The system's performance was assessed based on task completion time, the accuracy of robot movements, and participants' subjective feedback on the ease of interface use.

#### D. Data Analysis

Task execution times and robot movement accuracy were recorded for each participant and analyzed using descriptive statistics. Subjective opinions were collected using the System Usability Scale (SUS) method and analyzed through qualitative approaches. SUS provides a "quick and dirty", reliable tool for measuring the usability, it consists of a 10 item questionnaire with five response options for respondents; from Strongly agree to Strongly disagree [29]. The multimodal user interface was implemented and tested on a clinical robot within a simulated laboratory setting. System performance was assessed in terms of accuracy, speed, and user-friendliness.

### IV. RESULTS AND DISCUSSIONS

The multimodal user interface was implemented and tested on a test clinical robot in a laboratory environment. The system's performance was evaluated in terms of accuracy, speed, and user-friendliness. The tests were carried out with a prototype of a robotic arm manufactured with three stepper

motors and an Arduino Uno microcontroller. Fig. 9 shows the system server made up of a Raspberry Pi 4, glove with accelerometer, electronic components and the Virtual Reality headset.



Fig. 9. Electronic components of the developed system.

#### A. Voice Command Control

Voice control proved effective in maneuvering the robot and executing various commands. The accuracy of the voice recognition system was evaluated using a speech recognition rate metric, which measures the percentage of correctly recognized commands out of the total number of given commands. The speech recognition rate was 92%, indicating a high level of accuracy in recognizing voice commands.

#### B. Glove Control with Accelerometer

The accelerometer-equipped glove proved to be effective in capturing hand gestures and providing a natural way to control the robot. Fig. 10, 11, 12 displays the graph obtained by comparing accelerometer values along its 3 axes (X, Y, and Z) with the angles of rotation of the robot arm's corresponding 3 axes.

Fig. 10, 11, and 12 depict each of the three accelerometer axes positioned within the operator's gauntlet. The operator executed hand movements for a duration of one minute, yielding a total of 500 samples collected per accelerometer axis. These measurements correspond to the angular velocity ( $^{\circ}/s$ ) of motion recorded during the trials. Fig. 10 showcases the data acquired from the X-axis accelerometers, encompassing both the gauntlet and the robotic arm, while Y-axis data is presented in Fig. 11, and Z-axis data is delineated in Fig. 12. On average, a variation of 3.87% was observed, attributed to the motor configurations driven by the actuators.

#### C. Multimodal Control

The three control methods were amalgamated to forge a comprehensive multimodal user interface. Users were empowered to seamlessly switch between diverse control modes, tailoring their choice based on personal preferences and the specific task at hand. Empirical evidence substantiated the

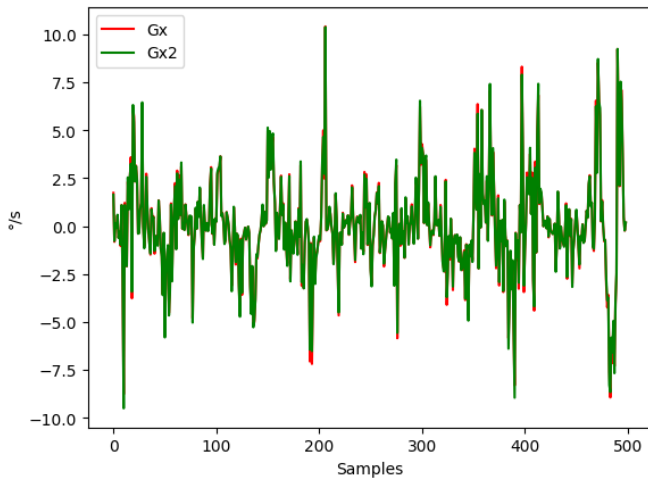


Fig. 10. Plot of accelerometer measurements and rotation of the robotic arm axes in the X-axis.

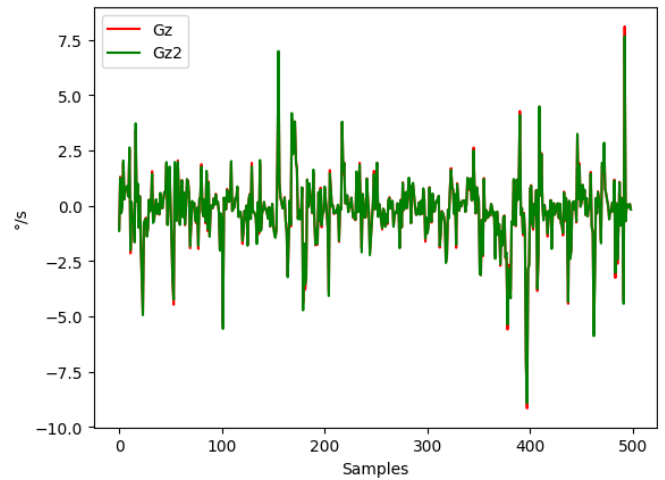


Fig. 12. Plot of accelerometer measurements and rotation of the robotic arm axes in the Z-axis.

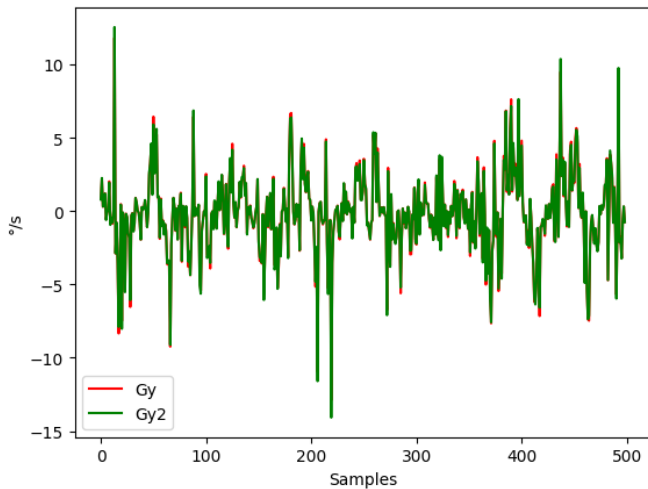


Fig. 11. Plot of accelerometer measurements and rotation of the robotic arm axes in the Y-axis.

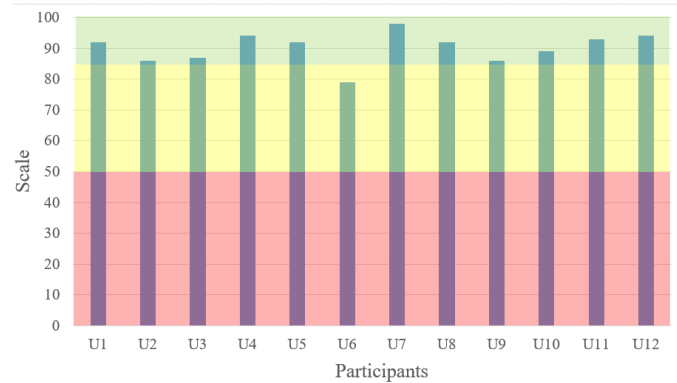


Fig. 13. Graph of results obtained from SUS measurement of the system.

superiority of the multimodal interface over singular control approaches, demonstrating that users adeptly transitioned between voice commands, hand gestures, and artificial vision control. This versatility imbued human-robot interaction with enhanced flexibility and intuitive fluidity. Moreover, the amalgamation of distinct modalities endowed a heightened precision of control, particularly advantageous for tasks necessitating meticulous accuracy, such as surgical procedures. Refer to Fig. 13 for a graphical representation of the values obtained through the evaluation of the user interface and the proposed system, utilizing the System Usability Scale method (SUS), a widely adopted metric for gauging the effectiveness of an interface for a given task. The color background of this graph shows three different scoring areas: light red for poor usability ( $SUS_{score} < 50$ ), light yellow for good usability ( $85 > SUS_{score} \geq 50$ ), and light green for excellent usability ( $SUS_{score} \geq 85$ ).

The obtained average value for the proposed interface was

$SUS_{score} = 90.2$  points, falling within the range indicative of commendable interfaces. These findings strongly indicate that the suggested user interface is exceptionally well-suited for orchestrating robotic arms within clinical scenarios. On the whole, the outcomes of this study strongly propose that the developed multimodal user interface holds substantial potential for enhancing the efficiency and efficacy of clinical robots within healthcare settings. The capability to govern the robot through voice commands, hand gestures, and artificial vision confers a heightened level of flexibility and intuitive interaction with the robotic system. This, in turn, stands to enhance patient outcomes and foster a higher adoption rate of the technology amongst healthcare professionals.

## V. CONCLUSIONS

In summary, a multimodal user interface has been introduced for touchless control of a clinical robot, seamlessly integrating voice commands, an accelerometer-equipped gauntlet, and display of the user interface on the virtual reality headset in real time. The outcomes derived from the conducted trials robustly suggest that the utilization of a multimodal interface holds the potential to enhance the efficiency and efficacy of clinical robots within healthcare environments, as



evidenced by the notable 90.2 point outcome on the SUS scale. The capacity to manipulate a robotic arm through the fusion of voice commands, hand gestures, and artificial vision engenders a more adaptable and intuitive means of interacting with the arm, a facet that has the potential to enhance patient outcomes and bolster the technology's embrace amongst healthcare professionals. Future endeavors will be concentrated on refining the interface and appraising its effectiveness within clinical environments, involving real patients. Additionally, the incorporation of other modalities, such as haptic feedback and augmented reality, could be explored to further heighten user experience and system performance.

#### ACKNOWLEDGMENT

The authors would like to thanks Universidad Nacional de San Agustín de Arequipa.

#### REFERENCES

- [1] Giovanni Morone, Ilaria Cocchi, Stefano Paolucci and Marco Iosa. Robot-assisted therapy for arm recovery for stroke patients: state of the art and clinical implication. *Expert Review of Medical Devices* 2020, vol. 17, pp. 223-233.
- [2] Mahdiah Babaiasl, Seyyed Hamed Mahdioun, Poorya Jaryani and Mojtaba Yazdani. A review of technological and clinical aspects of robot-aided rehabilitation of upper-extremity after stroke. *Disability and Rehabilitation: Assistive Technology* 2015, vol. 11, pp. 263-280.
- [3] Rafael Verano M, Jose Caceres S, Abel Arenas H, Andres Montoya A, Joseph Guevara M, Jarelh Galdos B and Jesus Talavera S, "Development of a Low-Cost Teleoperated Explorer Robot (TXRob)" *International Journal of Advanced Computer Science and Applications(IJACSA)*, 13(7), 2022.
- [4] V. Duchaine and C. Gosselin. Safe, Stable and Intuitive Control for Physical Human-Robot Interaction. 2009 IEEE International Conference on Robotics and Automation, Kobe, Japan, 2009, pp. 3383-3388.
- [5] William Percy Cuno Zuniga, Jordan Jonathan Quispe Navarro, Juan Daniel Pocohuanca Diaz, Jesus Talavera S., Andres Montoya A. Design of a Terrain Mapping System for Low-cost Exploration Robots based on Stereo Vision. *Przeegląd Elektrotechniczny* 2023.
- [6] Maurtua I, Fernández I, Tellaeché A, et al. Natural multimodal communication for human-robot collaboration. *International Journal of Advanced Robotic Systems*. 2017;14(4).
- [7] Schreiter, J., Mielke, T., Schott, D. et al. A multimodal user interface for touchless control of robotic ultrasound. *Int J CARS* 18, 1429–1436 (2023).
- [8] Sagar V. Fegade, Ashish Chaturvedi, Mukta Agarwal. Voice Recognition Technology : A Review. *International Journal of Advanced Research in Science, Communication and Technology*. 2021, vol. 8, no. 1.
- [9] Cristina Bayón, Rafael Raya, Sergio Lerma Lara, Oscar Ramirez, Ignacio Serrano, Eduardo Rocon. Robotic Therapies for Children with Cerebral Palsy: A Systematic Review. *Translational Biomedicine*. 2016, vol. 7, no. 1.
- [10] Jung, S.W., Park, M.Y., Park, I.M., Jung, Y.K., Shin, H.B. A Robust Control of Intelligent Mobile Robot Based on Voice Command. *Intelligent Robotics and Applications. ICIRA 2013. Lecture Notes in Computer Science*, vol 8102. Springer, Berlin, Heidelberg.
- [11] I. Cohen, B. Berdugo. Microphone array post-filtering for non-stationary noise suppression. *Proceedings IEEE International Conference on Acoustics, Speech, and Signal Processing*. 2002, pp. 901–904.
- [12] Jon Barker, Martin Cooke, Phil Green. Robust ASR based on clean speech models: An evaluation of missing data techniques for connected digit recognition in noise. *7th European Conference on Speech Communication and Technology*. 2001, pp. 213-217.
- [13] Philippe Renevey, Rolf Vetter, Jens Krauss. Robust speech recognition using missing feature theory and vector quantization. *7th European Conference on Speech Communication and Technology*. 2001, pp. 1107-1110.
- [14] S. Yamamoto, Kazuhiro Nakadai, Hiroshi Tsujino, Hiroshi Okuno. Assessment of general applicability of robot audition system by recognizing three simultaneous speeches. *International Conference on Intelligent Robots and Systems*. 2004, vol. 3, pp. 2111-2116.
- [15] Jean-Marc Valin, Jean Rouat, François Michaud. Enhanced Robot Audition Based on Microphone Array Source Separation with Post-Filter. 2004, vol. 3. pp. 2123-2128.
- [16] Xing Li. Human-robot interaction based on gesture and movement recognition. *Signal Processing: Image Communication*, vol. 81, 2020.
- [17] Ahmad Athif Mohd Faudzi, Muaammar Hadi Kuzman Ali, M. Asyraf Azman, Zool Hilmi Ismail. Real-time Hand Gestures System for Mobile Robots Control. *Procedia Engineering*, vol. 41, pp. 798-804, 2012.
- [18] Malima, A., E. Ozgur, et al. A Fast Algorithm for Vision-Based Hand Gesture Recognition for Robot Control. *Signal Processing and Communications Applications*, 2006, pp. 1-4.
- [19] Siddharth S. Rautaray, Anupam Agrawal. Real Time Gesture Recognition System for Interaction in Dynamic Environment. *Procedia Technology*. 2012, vol. 4, pp. 595-599.
- [20] Correa, M., Ruiz-del-Solar, J., Verschae, R., Lee-Ferng, J., Castillo, N. Real-Time Hand Gesture Recognition for Human Robot Interaction. *Lecture Notes in Computer Science 2009*, vol 5949. Springer, Berlin, Heidelberg.
- [21] Sichao Liu, Lihui Wang, Xi Vincent Wang. Symbiotic human-robot collaboration: multimodal control using function blocks. *Procedia CIRP*. 2020, vol. 93, pp. 1188-1193.
- [22] Alejandro Jaimes, Nicu Sebe. Multimodal human-computer interaction: A survey. *Computer Vision and Image Understanding*. 2007, vol. 108, pp. 116-134.
- [23] Lunghi, G.; Marin, R.; Di Castro, M.; Masi, A.; Sanz, P.J. Multimodal Human-Robot Interface for Accessible Remote Robotic Interventions in Hazardous Environments. *IEEE Access*. 2019, vol. 7, pp. 127290–127319.
- [24] S. Poirier, F. Routhier, A. Campeau-Lecours. Voice Control Interface Prototype for Assistive Robots for People Living with Upper Limb Disabilities. *International Conference on Rehabilitation Robotics*, Toronto, ON, Canada. 2019, pp. 46-52.
- [25] Morgan AA, Abdi J, Syed MAQ, Kohen GE, Barlow P, Vizcaychipi MP. Robots in Healthcare: a Scoping Review. *Curr Robot Rep*. 2022, 3(4):271-280.
- [26] Zhang J, Wang B, Zhang C, Xiao Y and Wang MY (2019) An EEG/EMG/EOG-Based Multimodal Human-Machine Interface to Real-Time Control of a Soft Robot Hand. *Front. Neurobot*. 13:7.
- [27] M.H. Korayem, M.A. Madihi, V. Vahidifar. Controlling surgical robot arm using leap motion controller with Kalman filter. *Measurement*, vol. 178, pp. 109372, 2021.
- [28] Ferre M, Macias-Guarasa J, Aracil R, Barrientos A. Voice command generation for teleoperated robot systems. 1998.
- [29] Brooke J. SUS: A quick and dirty usability scale. *Usability Evaluation in Industry*. 1996, pp. 189-194.

# Dual-Level Blind Omnidirectional Image Quality Assessment Network Based on Human Visual Perception

Deyang Liu<sup>1</sup>, Lu Zhang<sup>2</sup>, Lifei Wan<sup>3</sup>, Wei Yao<sup>4</sup>, Jian Ma<sup>5\*</sup>, Youzhi Zhang<sup>6</sup>

School of Computer and Information, Anqing Normal University, Anqing, 246000, China<sup>1,2,3,6</sup>

School of Teacher Education, Anqing Normal University, Anqing, 246000, China<sup>4</sup>

School of Computer Science, Fudan University, Shanghai 200433, China<sup>5</sup>

**Abstract**—With the rapid development of virtual reality (VR) technology, a large number of omnidirectional images (OIs) with uncertain quality are flooding into the internet. As a result, Blind Omnidirectional Image Quality Assessment (BOIQA) has become increasingly urgent. The existing solutions mainly focus on manually or automatically extracting high-level features from OIs, which overlook the important guiding role of human visual perception in this immersive experience. To address this issue, a dual-level network based on human visual perception is developed in this paper for BOIQA. Firstly, a human attention branch is proposed, in which the transformer-based model can efficiently represent attentional features of the human eye within a multi-distance perception image pyramid of viewport. Then, inspired by the hierarchical perception of human visual system, a multi-scale perception branch is designed, in which hierarchical features of six orientational viewports are considered and obtained by a residual network in parallel. Additionally, the correlation features among viewports are investigated to assist the multi-viewport feature fusion, in which the feature maps extracted from different viewports are further measured for their similarity and correlation by the attention-based module. Finally, the output values from both branches are regressed by fully connected layer to derive the final predicted quality score. Comprehensive experiments on two public datasets demonstrate the significant superiority of the proposed method.

**Keywords**—Omnidirectional image quality assessment; dual-level network; human visual perception; human attention; multi-scale

## I. INTRODUCTION

Virtual reality (VR), as the most popular immersive multimedia, can offer a unique 360-degree visual experience which sets it apart from traditional two-dimensional (2D) formats. Users can explore omnidirectional images (OIs) by wearing VR devices such as head-mounted displays (HMDs). However, the qualities of OIs are degraded during the processes such as stitching, projecting, encoding, and transmitting, which further influence the user experiences, even cause motion sickness. Therefore, the quality evaluation of OIs plays a significant role in guiding OI processing and ensuring a high quality of experience.

In the past few years, many objective OIQA methods have been proposed, including full-reference (FR) type and blind/no-reference (B/NR) type. For FR type, the peak-signal-to-noise ratio (PSNR) and the structural similarity index measure (SSIM) [1] are respectively adopted for OIQA [2, 3, 4, 5, 6]. However, obtaining the undistorted OIs is challenging due to the complexity of image processing, making FR-OIQA

challenging in practical applications. Therefore, it is crucial to develop blind/no-reference omnidirectional image quality assessment (BOIQA/NR-OIQA) methods that can evaluate the quality of OIs without reference images. Regarding the NR-OIQA type, many approaches [7, 8, 9, 10] commonly involve filtering to analyze the frequency domain information or natural scene statistics (NSS) to find statistical regularities in OIs. However, the manual feature designing is challenging [11], which limits the robustness of those methods. To relieve this issue, many data-driven approaches are proposed, which are able to learn inherent relationships between the predicted values and the ground-truth labels. These methods typically consist of two steps: feature extraction and quality regression. Specifically, Convolutional Neural Networks (CNNs) are firstly used to extract high-level features from OIs. Then, fully connected layers are employed for regression to obtain the predicted quality scores. However, most data-driven solutions are directly transferred from 2D IQA methods, in which the features of OIs are extracted in EquiRectangular Projection (ERP) format. Moreover, those approaches do not consider the human visual perception during OIQA. Although several approaches [12, 13, 14, 15] try to extract the viewport (VP) images from OIs to replace the ERP as the inputs, the human visual perception is still under-explored.

Generally, people tend to pay more attention to some contents of interest rather than the entire VP with HMD. This means that the regions of interest in VP are more likely to contribute to the quality rating [15]. Furthermore, the objects in nature are usually captured by the human eyes at various scales [16], which means that the human visual perception of an OI is formed through multiple views from different directions at various viewing distances.

Based on the above analysis, we can conclude that the quality of immersive media experiences is more susceptible to subjective visual perception by humans. However, recent works on OIQA primarily analyze images and overlook the active nature of human visual perception in this process. To address this gap and further to enhance the OIQA performance, this paper proposes a dual-level BOIQA network based on human visual perception. The proposed method tries to explore the human visual perception from two aspects including the human attention and the multi-scale perception. Specifically, the proposed method is a dual-level model, which is composed of three parts: human attention branch (HAB), multiscale perception branch (MPB) and quality regression (QR). For the

HAB, to emphasize the regions of interest, an improved Vision Transformer (ViT) [17] is integrated with the residual CNNs, which enables the proposed network to capture attention-based features within the VP images without disrupting the hierarchical perception. In HAB, the CNNs are responsible for obtaining high-level feature maps of each VP image, while the ViT calculates the attention weights. Furthermore, in order to explore more information within the VP region, we also introduce an image pyramid to represent different viewing distances of each VP in HAB. Regarding the MPB, we first establish a parallel structure to extract multi-scale information from each VP in cubemap projection (CMP) format. Then, to explore the content correlations between VPs at different positions in an OI, we develop a correlation feature fusion module to establish the long-range dependencies among VPs. Finally, the obtained dual-level perception features are regressed through the QR module to predict the final quality scores. Extensive experimental results have validated the effectiveness of the proposed approach. The contributions of the proposed method are listed as follows:

- We propose a BOIQA network based on the human visual perception, in which the region of interest can better be highlighted in a VP region and the multi-scale information can be obtained from low-level to high-level based on multiple views.
- We establish the multiple viewing distances image pyramid of the front VP and obtain the attention-based features from it to explore more information within the VP region. Moreover, we fuse the multi-scale features extracted from each VP in CMP and the obtained attention-based features to explore the content correlations between VPs.
- Comparisons with the state-of-the-art metrics on two public databases demonstrate the strength of our method.

## II. RELATED WORKS

Generally, OIQA methods can be classified into two categories: subjective methods and objective methods. Subjective OIQA method involves participants directly providing subjective quality scores for the OIs they view in an HMDs. However, it is time-consuming and impractical for batch applications. By contrast, objective OIQA method is more suitable for practical production applications. The objective OIQA method can be further divided into two categories: traditional OIQA metrics and deep learning-based OIQA metrics. This section will emphatically review the objective OIQA methods.

### A. Traditional OIQA Metrics

Many works have extended the traditional common used IQA metrics to OIQA. For example, the evaluation schemes based on PSNR transfer the calculation from planar format to spherical format while still inheriting the main idea of per-pixel comparison in PSNR. Moreover, the evaluation schemes based on SSIM mainly focus on the ERP format of panoramic images and analyze metrics such as sharpness, contrast, and brightness. In [18], statistical characteristics of panoramic images were obtained using the adjacent pixels correlation (APC) features and blind quality prediction of panoramic images was then performed using support vector regression (SVR).

The methods that use the ERP as the evaluation basis are mostly borrowed directly from 2D-IQA and have made corresponding improvements for panoramic images. However, they still overlook the unique media characteristics of panoramic images and the geometric distortions present in ERP. Recent works have focused on extracting natural statistical information from other representations of panoramas. Zheng et al. [19] firstly converted the panoramic image from the ERP format to a segmented spherical projection format. They then utilized a heat map as a weighting factor to perceive features in both the two-level and equatorial regions. Zhou et al. [9] achieved panoramic image quality assessment score by analyzing multi-frequency information and statistically evaluating the local and global naturalness presented in both ERP and VP formats. Jiang et al. [8] explored the color information of each VP image unit in the rotated Cubemap Projection (CMP) format through tensor decomposition and piecewise exponential fitting. The above-mentioned works achieved satisfactory performance results by designing hand-crafted features through techniques such as machine learning. However, these manual features based approaches are evidently cumbersome and not easily comprehensive, which reduces the robustness of the proposed method.

### B. Deep Learning-based OIQA Metrics

Deep learning-based OIQA approaches benefit from powerful model architectures that can capture more quality-relevant features within the images. Thanks to the guidance of large amounts of labeled data, this kind of methods often outperforms traditional methods. In [20, 21], Kim et al. proposed an adversarial learning-based human perception guider, which improves the prediction capability of deep learning models for quality scores by enhancing the human perception guider's discriminative ability for predicted scores and subjective quality score labels.

Although the aforementioned methods have achieved satisfactory results, they have not considered the differences between immersive media experience and traditional planar images perception. This restricts the feature representation capability of deep models. To address this issue, recent VP-based end-to-end models have been developed to accurately simulate the scene content that can be perceived by the human eye while viewing panoramic images at a moment. Considering the limited field of view of the human eye in head-mounted devices, Li et al. [12] firstly proposed a VP-based assessment scheme and combined it with CNNs for feature extraction. Sun et al. [13] proposed a multi-stream network that utilized the modified ResNet-34 to extract features from each VP in the rotated CMP format. Xu et al. [14] proposed a solution with local and global branches. The local branch utilized ResNet-18 to simultaneously extract internal features from multiple VP images and established connections between them using graph convolution. The global branch extracted feature information from the panoramic ERP format using the VGG [22] network.

These deep learning-based models possess powerful feature extraction capabilities and quality score fitting abilities. However, there is still significant room for improvement in terms of their consistency with the HVS. Therefore, in this paper, we draw inspiration from human visual perception and develop an end-to-end model to investigate the impact of

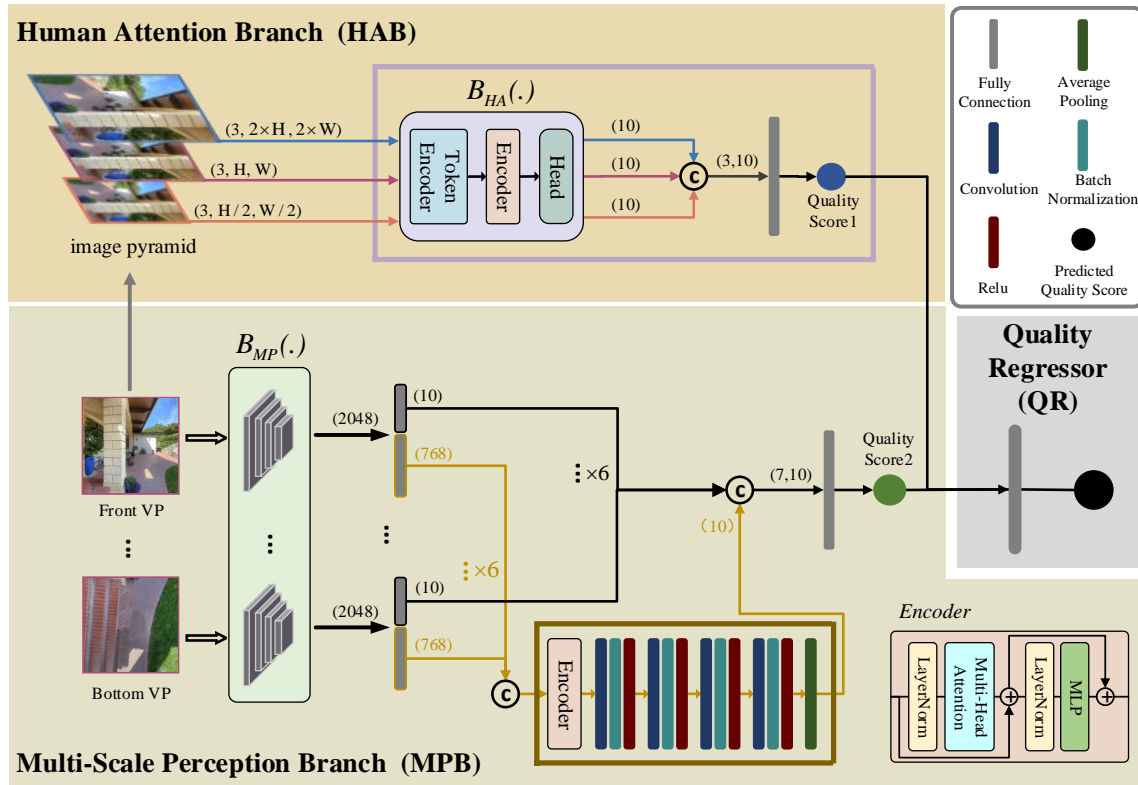


Fig. 1. The overall framework structure of our dual-level network

simulating human attention [23] and multi-scale [24] perception on panoramic quality assessment.

### III. PROPOSED METHOD

The proposed dual-level BOIQA network contains three modules, namely human attention branch (HAB), multi-scale perception branch (MPB) and quality regression (QR). The overall framework structure is illustrated in Fig. 1. The HAB focusses on extracting the high-level information from a multiple viewing distances image pyramid based on attention mechanism. The MPB aims to explore multi-scale perception features of each VP and explore the correlation information among those VPs. The QR is utilized to predict perceptual quality scores.

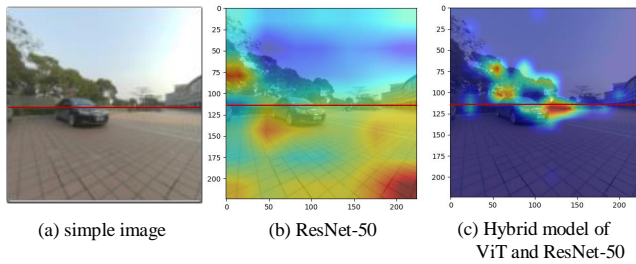


Fig. 2. the visualization of attention from the output features to a VP example from CVIQD database. The solid red line shows the position of the equator. (a) the simple image; (b) the learned feature maps of (a) only with ResNet-50; (c) the learned feature maps of (a) with the hybrid model of ResNet-50 and ViT.

#### A. Human Attention Network

Fig. 2 shows the visualization result of the attention weights on a VP example under different OIQA models. It is evident that the hybrid model combining CNNs and ViT pays more attention to the equatorial region and salient objects compared to a pure CNNs model. This aligns well with the attention habits of the HVS. Based on the above analysis, to obtain information that better aligns with human visual attention characteristics, the HAB branch is designed to extract internal features of VP images based on attention mechanisms. To further explore the comprehensive perception of the VP under different viewing distances, we also introduced a multiple viewing distances image pyramid of the front VP as the input of the HAB branch, which is shown in Fig. 1.

For the front VP initialized with a resolution of  $H \times W$  through the center cropping operation, we increase the center cropping resolution to  $2H \times 2W$  to represent a larger field of view with a longer distance. Similarly, we decrease the center cropping resolution to  $\frac{H}{2} \times \frac{W}{2}$  to represent a smaller field of view with a closer distance. Therefore, the input image pyramid  $V_f^l, V_f^o, V_f^s$  of the front VP can be established with a multiple distances representation. Specifically, the  $V_f^o$  represents the VP in its original resolution,  $V_f^l$  represents a version with a higher resolution, and  $V_f^s$  represents a version with a lower resolution.

To fully explore the information based on human visual attention from this VP pyramid, we integrate the ResNet-50 and an improved ViT as the backbone for feature extraction. Specifically, each layer of the pyramid is firstly fed into the ResNet-50 in parallel. The semantic features of each view can

be obtained and then being converted into token forms. In our method, the improved ViT network consists of two stages. The first stage tries to compute the attention weight among tokens within each view by multi head attention (MHA). The second stage is used to further adjust the dimensionality of the obtained feature maps based on view-content attention through the Head module which is composed of fully connected layers. Finally, we fuse the extracted high-level features of each view at different distances, and obtain the quality score of this branch with a regression operation. This process can be expressed as:

$$\begin{cases} F_f^l, F_f^o, F_f^s = B_{HA}(V_f^l, V_f^o, V_f^s), \\ Q_1 = Linear(Cat(F_f^l, F_f^o, F_f^s)), \end{cases} \quad (1)$$

where  $B_{HA}(\cdot)$  represents the feature extraction network of the HAB,  $F_f^l, F_f^o, F_f^s$  separately represent the extracted features from the image pyramid under different distances. Each extracted features of the pyramid has a dimensionality of 10.  $Cat(\cdot)$  and  $Linear(\cdot)$  respectively donate the concatenate operation and the fully connected layer.  $Q_1$  is the obtain predicted quality score of this process.

### B. Multi-Scale Perception Network

In general, the user's comprehensive quality perception of a panoramic image is influenced by multiple VPs at different positions. It is necessary to perform multi-scale quality perception across various positions in the panoramic image and explore the correlation information between these VPs in terms of both location and content.

In this work, we have established a multi-scale perceptual branch as an auxiliary branch. Firstly, a group of VP images  $V_u, V_d, V_l, V_r, V_f, V_b$  are achieved from a panoramic image at six directions (up, down, left, right, front, and back). As mentioned above, human visual perception is a hierarchical process that involves perceiving texture, contours, and high-dimensional semantics. Therefore, in this branch, ResNet-50 with residual structure is adopted as the backbone for feature extraction of each VP. The residual network is capable of capturing multi-scale perceptual features from low-level to high-level in each directional VP, which aligns well with the multi-scale perception of HVS. This process can be represented as:

$$F_i^m = B_{MP}(V_i), i \in \{u, d, l, r, f, b\}, \quad (2)$$

For each VP image  $V_i$ , multi-scale perceptual features  $F_i^m$  are simultaneously extracted through the feature extraction network.  $B_{MP}$  represents the backbone of this branch. It is worth noting that the multi-scale features obtained here have a dimensionality of 2048.

After obtaining the multi-scale features  $F_i^m$  obtained from multiple VP images, most methods propose to concatenate those high-dimensional features and perform quality regression. Conversely, in order to further capture the inter-viewpoint correlation information, we apply a fully connected layer to convert the multi-scale features corresponding to each VP into tokens with a dimensionality of 768. Subsequently, we perform element-wise multiplication operations based on attention mechanism among those tokens to obtain the correlational features. The specific formula representation is as follows.

$$\begin{cases} T_i^m = Linear(F_i^m) \\ F^c = E_{MP}(T_i^m) \end{cases}, i \in \{u, d, l, r, f, b\}, \quad (3)$$

Here,  $Linear(\cdot)$  represents the fully connected operation,  $T_i^m$  denotes the token corresponding to each VP after undergoing this fully connected operation.  $E_{MP}$  represents the Encoder network based on multi-head attention, and  $F^c$  represents the final correlation feature map among those VPs.

In order to further fuse the obtained correlation features and consider the spatial relationship between each VP, we further introduce a correlation feature fusion module. This module consists of four convolutional blocks and an average-pooling. Each convolutional block includes a convolution (Conv) layer, a batch normalization (BN) layer, and a Rectified Linear Unit (ReLU) activation. The convolutional operation calculates the internal correlations of the feature map  $F^c$  using a  $3 \times 3$  receptive field, integrating the content-based correlation information between different VPs. This locally nested convolutional structure also helps compensate for the positional correlation between VPs that may be overlooked in the previous computations. Finally, an average-pooling operation is applied to obtain the fused correlation features through regression. The specific process is illustrated by the following equation.

$$F^{c'} = C_{MP}(F^c), \quad (4)$$

where  $C_{MP}$  represents the correlation feature fusion module,  $F^{c'}$  is the achieved fused correlation features, whose dimensionality is 10.

We perform final feature regression on the multi-scale information obtained from each VP and the corresponding fusion information between them. Specifically, we first adjust each multi-scale feature map  $F_i^m$  to dimensionality 10. Then, we concatenate those adjusted feature maps with the multi-scale feature  $F^{c'}$ . Next, a fully connected operation is applied to the concatenated feature map for feature regression. This process can be described as:

$$\begin{cases} F_i^{m'} = Linear(F_i^m) \\ Q_2 = Linear(Cat(F_i^{m'}, F^{c'})) \end{cases} \quad i \in \{u, d, l, r, f, b\} \quad (5)$$

where  $F_i^{m'}$  represents the multi-scale features of each VP after dimension adjustment.  $Linear(\cdot)$  and  $Cat(\cdot)$  represent the fully connected operation and concatenation operation, respectively.  $Q_2$  denotes the quality score obtained from the final regression of this multi-scale perception branch.

### C. Quality Regressor

The quality regressor consists of two steps. Firstly, we conduct concatenation operation of HAB and MPB. Afterwards, the predicted score is obtained by the final layer of fully connected. The training loss is described as follows:

$$\begin{cases} Q = Linear(Cat(Q_1, Q_2)) \\ L = |Q - MOS|^2, \end{cases} \quad (6)$$

where  $Q$  is the final predicted quality score. The  $MOS$  is the ground-truth label of OI, which also means the subjective

TABLE I. OVERALL PERFORMANCE COMPARISONS ON THE OIQA AND CVIQD DATABASES. THE BEST RESULTS ARE DENOTED IN BOLD.

Type	Database	OIQA			CVIQD		
	Methods	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE
FR	PSNR	0.5812	0.5226	1.7005	0.7008	0.6239	9.9599
	S-PSNR	0.5997	0.5399	1.6721	0.7083	0.6449	9.8564
	WS-PSNR	0.5819	0.5263	1.6994	0.6729	0.6107	10.3283
	CPP-PSNR	0.5683	0.5149	1.7193	0.6871	0.6265	10.1448
	SSIM	0.8718	0.8588	1.0238	0.9002	0.8842	6.0793
	MS-SSIM	0.7710	0.7379	1.3308	0.8521	0.8222	7.3072
	FSIM	0.9014	0.8938	0.9047	0.9340	0.9152	4.9864
	DeepQA	0.9044	0.8973	0.8914	0.9375	0.9292	4.8574
NR	BRISQUE	0.8424	0.8331	1.1261	0.8376	0.8180	7.6271
	BMPRI	0.6503	0.6238	1.5874	0.7919	0.7470	8.5258
	DB-CNN	0.8852	0.8653	0.9717	0.9356	0.9308	4.9311
	MC360IQA	0.9267	0.9139	0.7854	0.9429	0.9428	4.6506
	VGCN	0.9584	0.9515	0.5967	0.9651	0.9639	3.6573
	Ours	<b>0.9598</b>	<b>0.9530</b>	<b>0.5862</b>	<b>0.9680</b>	<b>0.9664</b>	<b>3.5014</b>

quality score. The  $L$  represents the loss between  $S$  and  $MOS$ , the  $|\cdot|$  represents absolute value operation.

#### IV. EXPERIMENTAL RESULTS

Our experiment uses two popular public datasets, namely CVIQD [25] and OIQA[26]. They both include 16 original panoramic images with different types and degrees of distortion. The former includes 528 compressed images generated by JPEG, H.264/AVC and H.265/HEVC standards, and the subjective score label of it ranges from 1 to 100. The latter contains 320 distorted images generated by four distortion types: JPEG compression (JPEG), JPEG2000 compression (JP2K), Gaussian blur (GB) and Gaussian white noise (GN), and the subjective ground-truth label of it ranges from 1 to 10.

##### A. Experimental Settings

Our experiments were conducted with 11<sup>th</sup> Gen Intel(R) Core(TM) CPU i7-11700F @ 2.50GHz, 16 GB RAM, NVIDIA RTX 3060. The batch size was set to 4 and the learning strategy was RMSprop [27] whose learning rate is initialized to 0.0001. The rotation angle for the rotated CMP was fixed to 4 serving as data augmentation and the VP image resolution  $H \times W$  is set to  $256 \times 256$ . Each database is split into training and testing sets according to the standard ratio of 8:2. This means that the distorted images corresponding to 3 reference images are randomly selected as testing set and the remaining are regarded as the training set. During the training phase, we use the pretrain results of ImageNet to the HAB and the MPB's backbone. By transferring the model training parameters from a large dataset to our task-specific dataset, we can achieve significant benefits. For the Backbone of HAB, the number of the MHA is set to 8 and the number of encoder blocks is set to 11. Finally, we adopt three standard assessment methods: Pearsons linear correlation coefficient (PLCC), Spearman's rank order correlation coefficient (SROCC) and root mean squared error (RMSE) to assess the model performances. The former two respectively evaluate the prediction results based on rank correlation and linear correlation. A value closer to 1 indicates a better

prediction result. The latter measures the discrepancy between the predicted and ground-truth values, with a value closer to 0 indicating a better prediction result. We also used a five-parameter logistic function to fit the predicted quality scores and the ground-truth labels:

$$y = \beta_1 \left( \frac{1}{2} - \frac{1}{1 + \exp(\beta_2(x - \beta_3))} \right) + \beta_4 x + \beta_5, \quad (7)$$

where  $x$  refers to the predicted quality score and  $y$  represents the mapped score.  $\beta_1$  to  $\beta_5$  are five parameters.

##### B. Performance Evaluation

1) *Comparison Metrics*: In order to illustrate the effectiveness of our model, the comparison algorithms includes FR and NR OIQA metrics. The FR-OIQA metrics include PSNR, S-PSNR [2], WS-PSNR [3], CPP-PSNR [4], SSIM [1], MS-SSIM [5], FSIM [6] and DeepIQA [28]. The NR-OIQA contain BRISQUE [29], BMPRI [30], DB-CNN [31], MC360IQA [13], DDA-BOIQA [32] and VGCN [14].

The performance comparison results on the OIQA and CVIQD datasets are shown in Table I. Among these FR-OIQA methods, the PSNR-related algorithms which have weaker correlation with the HVS exhibit poorer performance compared to these state-of-the-art objective algorithms. It is a breakthrough that the SSIM takes into account the brightness, contrast, and structural features associated with the HVS. However, the evaluation results are still limited and the performances are inferior to deep learning-based FR-OIQA methods. The reason lies in that deep learning-based methods directly consider the internal relationship between images and subject scores, while other methods mainly focus on one or two features of the OI. For NR methods, these algorithms generally outperform FR-OIQA algorithms. BRISQUE, BMPRI, and DB-CNN are implemented for OIQA specifically targeting ERP format of OIs. Specifically, BRISQUE and BMPRI are implemented based on handcrafted feature designs, while DB-CNN is implemented based on a data-driven model. Furthermore, the MC360IQA and VGCN models consider the VP images into their CNN

TABLE II. PERFORMANCE COMPARISON ON OIQA DATABASE. THE BEST RESULT IS ANNOTATED WITH BOLD, AND THE SECOND-BEST RESULT IS ANNOTATED WITH UNDERLINE.

	JPEG			JP2K			WN			BLUR			
	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE	
FR	PSNR	0.6941	0.7060	1.6141	0.8632	0.7821	1.1316	0.9547	0.9500	0.5370	0.9282	0.7417	0.8299
	S-PSNR	0.6911	0.6148	1.6205	0.9205	0.7250	0.8757	0.9503	0.9357	0.5620	0.8282	0.7525	1.0910
	WS-PSNR	0.7133	0.6792	1.5713	0.9344	0.7500	0.9128	0.9626	0.9500	0.4890	0.8190	0.7668	1.1172
	CPP-PSNR	0.6153	0.5362	1.7693	0.8971	0.7250	0.9904	0.9276	0.9143	0.6739	0.7969	0.7185	1.1728
	SSIM	0.9077	0.9008	0.9406	0.9783	0.9679	0.4643	0.8828	0.8607	0.8474	<b>0.9926</b>	<u>0.9777</u>	<u>0.2358</u>
	MS-SSIM	0.9102	0.8937	0.9288	0.9492	0.9250	0.7052	0.9691	0.9571	0.4452	0.9251	0.8990	0.7374
	FSIM	0.8938	0.8490	1.0057	0.9699	0.9643	0.5454	0.9170	0.8893	0.7197	<u>0.9914</u>	<b>0.9902</b>	<b>0.2544</b>
	DeepQA	0.8301	0.8150	1.2506	<b>0.9905</b>	<b>0.9893</b>	<b>0.3082</b>	0.9709	<b>0.9857</b>	0.4317	0.9623	0.9473	0.5283
NR	BRISQUE	0.9160	0.9392	0.8992	0.7397	0.6750	1.5082	0.9818	0.9750	0.3427	0.8663	0.8508	0.9697
	BMPRI	0.9361	0.8954	0.7886	0.8322	0.8214	1.2428	0.9673	<u>0.9821</u>	0.4572	0.5199	0.3807	1.6584
	DB-CNN	0.8413	0.7346	1.2118	0.9755	0.9607	0.4935	0.9772	0.9786	0.3832	0.9536	0.8865	0.5875
	MC360IQA	0.9459	0.9008	0.7272	0.9165	0.9036	0.8966	0.9718	0.9464	0.4251	0.9526	0.9580	0.5907
	VGCN	<b>0.9540</b>	<b>0.9294</b>	<b>0.6720</b>	0.9771	0.9464	0.4772	<u>0.9811</u>	0.9750	<u>0.3493</u>	0.9852	0.9651	0.3327
	Ours	<u>0.9475</u>	<u>0.9133</u>	<u>0.7167</u>	<u>0.9885</u>	<u>0.9821</u>	<u>0.3390</u>	<b>0.9888</b>	0.9714	<b>0.2690</b>	0.9859	<u>0.9777</u>	0.3251

model, resulting in significant performance improvements. It is because the VP images are similar to the perception of human eyes. Our algorithm exhibits significantly higher performance compared to most deep learning based algorithms in terms of accuracy and monotonicity on those two databases. It is evident that the proposed dual-level network based on human visual perception is more consistent with the subject quality perception.

### 2) Performance Validity of Individual Distortion Types:

As illustrated in Table II and Table III, we also conducted comparative experiments of individual distortion types on OIQA and CVIQD. In general, our algorithm exhibits the best comprehensive performance for most of the distortions. The scatter plots in Fig. 3 and Fig. 4 depict the correlation between MOS and the predictions for individual distortion types on the two databases. These plots provide additional evidence to support the superiority of our approach. Specifically, our algorithm achieves top performance in both WN and AVC distortion types, and it closely follows the top-performing algorithm in JPEG and JP2K distortions. For example, as shown in Fig. 4, our algorithm exhibits an SROCC value in JPEG that is only 0.0161 lower than the top-performing VGCN, and is only 0.0062 lower than DeepQA on the OIQA. This further demonstrates the strong robustness of our algorithm in compression distortion types. Additionally, in terms of HEVC distortion, our algorithm achieves the best PLCC and RMSE values in NR-OIQA. It is noteworthy that SSIM and FSIM in FR-OIQA actually achieved the best results in this distortion type. The reason is that HEVC distortion typically includes color inaccuracies or artifacts, while FSIM primarily measures quality degradation by assessing the similarity in luminance, contrast, and structural aspects between the reference and distorted images. Therefore, FSIM is more sensitive to color changes. Additionally, FSIM benefits from having a reference image for comparison, which enhances its ability to identify

artifacts such as pseudo-imaging. This also indicates the effectiveness of these schemes for a certain type of distortion.

3) *Ablation Study:* In order to further demonstrate the effectiveness of each module, we separately removed each component of our model to conduct ablation experiments on two datasets. The experimental results are presented in Table IV. We separately adopt the human attention Branch (HAB) and the multi-scale perception branch (MPB) to predict the perceptual quality based on human visual perception in the dual-level network. In this section, we compare the performance with or without these two branches to respectively demonstrate the validity of each branch. We can conclude that both branches have strong quality prediction capabilities. However, the dual-stream network proposed in this paper, which combines features from these two branches, exhibits superior quality perception abilities, particularly in improving the SROCC values. Moreover, the influence of HAB is more pronounced on OIQA, mainly due to the diverse resolutions of OIs in this dataset, which are effectively addressed by the image pyramid utilized in the HAB. In our implementation, attentional features are obtained from an image pyramid in HAB. It is necessary to investigate how feature extraction based on attention mechanism affects overall performance. Therefore, we replace the original backbone in HAB with ResNet-50 and test the performance of the overall architecture. The results show that the performance of the CNN-based backbone is inferior to the backbone used in this paper, which further demonstrate the necessity of considering attentional features in HAB. In addition, we also conducted an ablation study for the correlation feature fusion module of VPs in MPB. As compared to the original implementation, the results on both databases showed slight improvements, which further proves that there is contextual and positional correlation information between different viewports within a panoramic image.

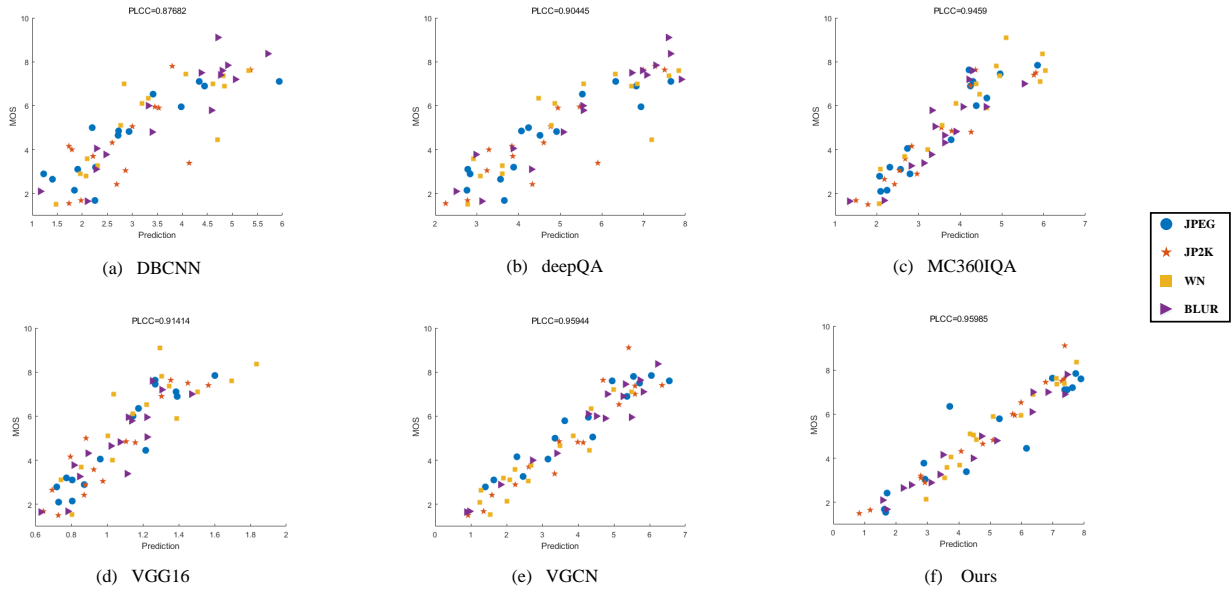


Fig. 3. Scatter plots of MOS values against predictions by OIQA metrics for individual distortion type on the testing set of OIQA Database.

TABLE III. PERFORMANCE COMPARISON ON CVIQD DATABASE. THE BEST RESULT IS ANNOTATED WITH BOLD, AND THE SECOND-BEST RESULT IS ANNOTATED WITH UNDERLINE.

		JPEG			AVC			HEVC		
		PLCC	SROCC	RMSE	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE
FR	PSNR	0.8682	0.6982	8.0429	0.6141	0.5802	10.5520	0.5982	0.5762	9.4697
	S-PSNR	0.8661	0.7172	8.1008	0.6307	0.6039	10.3760	0.6514	0.6150	8.9585
	WS-PSNR	0.8572	0.6848	8.3465	0.5702	0.5521	10.9841	0.5884	0.5642	9.5473
	CPP-PSNR	0.8585	0.7059	8.3109	0.6137	0.5872	10.5615	0.6160	0.5689	9.3009
	SSIM	0.9822	0.9582	3.0468	0.9303	0.9174	4.9029	<u>0.9436</u>	<u>0.9452</u>	<u>3.9097</u>
	MS-SSIM	0.9636	0.9047	4.3355	0.7960	0.7650	8.0924	0.8072	0.8011	6.9693
	FSIM	0.9839	0.9639	2.8928	0.9534	0.9439	4.0327	<b>0.9617</b>	<b>0.9532</b>	<b>3.2385</b>
	DeepQA	0.9526	0.9001	4.9290	0.9477	0.9375	4.2683	0.9221	0.9288	4.5694
NR	BRIAQUE	0.9464	0.9031	5.2442	0.7745	0.7714	8.4573	0.7548	0.7644	7.7455
	BMPRI	0.9874	0.9562	2.5597	0.7161	0.6731	9.3318	0.6154	0.6715	9.3071
	DB-CNN	0.9779	0.9576	3.3862	0.9564	0.9545	3.9063	0.8646	0.8693	5.9335
	MC360IQA	0.9698	0.9693	3.9517	0.9487	0.9569	4.2281	0.8976	0.9104	5.2557
	DDA-BOIQA	0.9570	0.9610	5.6010	0.9530	0.9490	3.8730	0.9290	0.9140	4.5250
	VGCN	<b>0.9894</b>	<u>0.9759</u>	<b>2.3590</b>	<u>0.9719</u>	<u>0.9659</u>	<u>3.1490</u>	0.9401	0.9432	4.0257
	Ours	<u>0.9878</u>	<b>0.9803</b>	<u>2.5285</u>	<b>0.9780</b>	<b>0.9796</b>	<b>2.7888</b>	0.9408	0.9405	4.0023

TABLE IV. ABLATION STUDY RESULTS FOR REMOVING EACH INDIVIDUAL BRANCH OR MODULE ON OIQA AND CVIQD.

Methods	OIQA			CVIQD		
	PLCC	SROCC	RMSE	PLCC	SROCC	RMSE
w/o HAB	0.9474	0.9414	0.6686	0.9623	0.9658	3.8786
w/o MPB	0.9572	0.9476	0.6049	0.9694	0.9627	3.4993
w/o attentional features in HAB	0.9482	0.9449	0.6639	0.9655	0.9650	3.7124
w/o correlation features in MPB	0.9569	0.9497	0.6072	0.9634	0.9632	3.8238
Ours	<b>0.9598</b>	<b>0.9530</b>	<b>0.5862</b>	<b>0.9680</b>	<b>0.9664</b>	<b>3.5014</b>



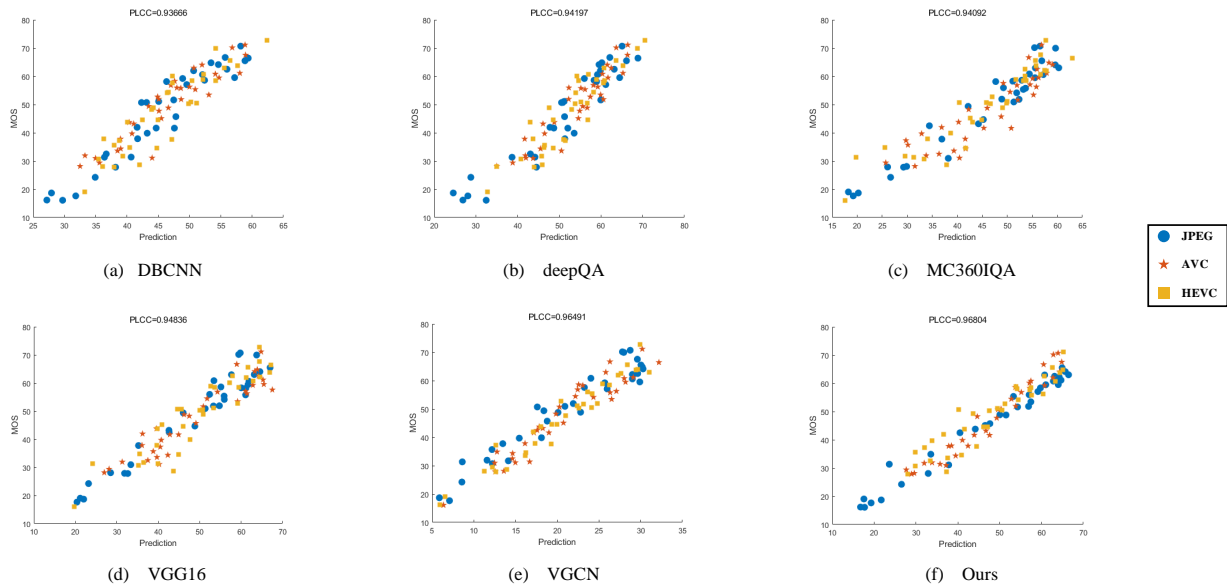


Fig. 4. Scatter plots of MOS values against predictions by OIQA metrics for individual distortion type on the testing set of CVIQD Database.

## V. CONCLUSION

In response to the fact that the perception of immersive media quality is more susceptible to subjective visual perception by the human eye, in this paper, we propose an innovative approach that integrates two characteristics of human visual perception, namely attentional perception and multi-scale perception, into the process of acquiring panoramic image features. Specifically, we propose a dual-level network based on human visual perception for blind omnidirectional image quality assessment. By transforming the front viewport image to an image pyramid with multiple viewing distances, the human attention branch is able to capture the high-level information based on attention mechanism. To obtain the features of different viewports from different position, we further establish a module to fuse their correlation information in the multi-scale feature perception branch after parallel extraction of their multi-scale features.

Experimental on two OIQA datasets show that our approach achieves the best performance, further validating the effectiveness of the human visual perception guidance. Of course, our work needs further in-depth research. Our approach only incorporates two essential aspects of human visual perception to assist the omnidirectional image quality assessment process. However, human visual perception is diverse, and the challenge lies in quantifying it effectively in a general end-to-end model. This will be the focus of our future research endeavors.

## ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China under Grant 62171002, 61906118, in part by STCSM under Grant SKLSFO2021-05, in part by University Discipline Top Talent Program of Anhui under Grant gxbjZD2022034, in part by Project on Anhui Provincial Natural Science Study by Colleges and Universities

under Grant 2022AH030106, in part by Key research projects in humanities and social sciences under Grant SK2019A0373, in part by Anhui educational science research project under Grant JK22007, and in part by China Postdoctoral Science Foundation under Grant 2022M710745.

## REFERENCES

- [1] Z. Wang et al. "Image quality assessment: from error visibility to structural similarity". In: *IEEE transactions on image processing* 13.4 (2004), pp. 600–612.
- [2] M. Yu, H. Lakshman, and B. Girod. "A framework to evaluate omnidirectional video coding schemes". In: *2015 IEEE international symposium on mixed and augmented reality*. IEEE. 2015, pp. 31–36.
- [3] Y. Sun, A. Lu, and L. Yu. "Weighted-to-spherically-uniform quality evaluation for omnidirectional video". In: *IEEE signal processing letters* 24.9 (2017), pp. 1408–1412.
- [4] V. Zakharchenko, K. P. Choi, and J. H. Park. "Quality metric for spherical panoramic video". In: *Optics and Photonics for Information Processing X*. Vol. 9970. SPIE. 2016, pp. 57–65.
- [5] Z. Wang, E. P. Simoncelli, and A. C. Bovik. "Multi-scale structural similarity for image quality assessment". In: *The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*. Vol. 2. Ieee. 2003, pp. 1398–1402.
- [6] L. Zhang et al. "FSIM: A feature similarity index for image quality assessment". In: *IEEE transactions on Image Processing* 20.8 (2011), pp. 2378–2386.
- [7] S. Ling, G. Cheung, and P. Le Callet. "No-reference quality assessment for stitched panoramic images using convolutional sparse coding and compound feature selection". In: *2018 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE. 2018, pp. 1–6.

- [8] H. Jiang et al. “Multi-Angle Projection Based Blind Omnidirectional Image Quality Assessment”. In: *IEEE Transactions on Circuits and Systems for Video Technology* 32.7 (2022), pp. 4211–4223. DOI: 10.1109/TCSVT.2021.3128014.
- [9] W. Zhou et al. “No-Reference Quality Assessment for 360-Degree Images by Analysis of Multifrequency Information and Local-Global Naturalness”. In: *IEEE Transactions on Circuits and Systems for Video Technology* 32.4 (2022), pp. 1778–1791. DOI: 10.1109/TCSVT.2021.3081182.
- [10] Y. Liu et al. “HVS-Based Perception-Driven No-Reference Omnidirectional Image Quality Assessment”. In: *IEEE Transactions on Instrumentation and Measurement* 72 (2023), pp. 1–11. DOI: 10.1109/TIM.2022.3232792.
- [11] S. Habib et al. “External Features-Based Approach to Date Grading and Analysis with Image Processing”. In: *Emerg. Sci. J* 6.4 (2022), pp. 694–704.
- [12] C. Li et al. “Viewport Proposal CNN for 360deo Quality Assessment”. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2019, pp. 10169–10178. DOI: 10.1109/CVPR.2019.01042.
- [13] W. Sun et al. “MC360IQA: The Multi-Channel CNN for Blind 360-Degree Image Quality Assessment”. In: *2019 IEEE International Symposium on Circuits and Systems (ISCAS)*. 2019, pp. 1–5. DOI: 10.1109/ISCAS.2019.8702664.
- [14] J. Xu, W. Zhou, and Z. Chen. “Blind Omnidirectional Image Quality Assessment With Viewport Oriented Graph Convolutional Networks”. In: *IEEE Transactions on Circuits and Systems for Video Technology* 31.5 (2021), pp. 1724–1737. DOI: 10.1109/TCSVT.2020.3015186.
- [15] A. Sendjasni and M.-C. Larabi. “SAL-360IQA: A Saliency Weighted Patch-Based CNN Model for 360-Degree Images Quality Assessment”. In: *2022 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*. IEEE. 2022, pp. 1–6.
- [16] Y. Lu et al. “Blind image quality assessment based on the multiscale and dual-domains features fusion”. In: *Concurrency and Computation: Practice and Experience* (2021), e6177.
- [17] A. Dosovitskiy et al. “An image is worth 16x16 words: Transformers for image recognition at scale”. In: *arXiv preprint arXiv:2010.11929* (2020).
- [18] W. Ding et al. “No-reference Panoramic Image Quality Assessment based on Adjacent Pixels Correlation”. In: *2021 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*. 2021, pp. 1–5. DOI: 10.1109/BMSB53066.2021.9547132.
- [19] X. Zheng et al. “Segmented Spherical Projection-Based Blind Omnidirectional Image Quality Assessment”. In: *IEEE Access* 8 (2020), pp. 31647–31659. DOI: 10.1109/ACCESS.2020.2972158.
- [20] H.-T. Lim, H. G. Kim, and Y. M. Ra. “VR IQA NET: Deep Virtual Reality Image Quality Assessment Using Adversarial Learning”. In: *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 2018, pp. 6737–6741. DOI: 10.1109/ICASSP.2018.8461317.
- [21] H. G. Kim, H.-T. Lim, and Y. M. Ro. “Deep Virtual Reality Image Quality Assessment With Human Perception Guider for Omnidirectional Image”. In: *IEEE Transactions on Circuits and Systems for Video Technology* 30.4 (2020), pp. 917–928. DOI: 10.1109/TCSVT.2019.2898732.
- [22] K. Simonyan and A. Zisserman. “Very deep convolutional networks for large-scale image recognition”. In: *arXiv preprint arXiv:1409.1556* (2014).
- [23] E. Z. Ye et al. “DeepImageTranslator V2: analysis of multimodal medical images using semantic segmentation maps generated through deep learning”. In: *bioRxiv* (2021), pp. 2021–10.
- [24] M. Jesmeen et al. “SleepCon: Sleeping Posture Recognition Model using Convolutional Neural Network”. In: *Emerging Science Journal* 7.1 (2022), pp. 50–59.
- [25] W. Sun et al. “A large-scale compressed 360-degree spherical image database: From subjective quality evaluation to objective model comparison”. In: *2018 IEEE 20th international workshop on multimedia signal processing (MMSP)*. IEEE. 2018, pp. 1–6.
- [26] H. Duan et al. “Perceptual Quality Assessment of Omnidirectional Images”. In: *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*. 2018, pp. 1–5. DOI: 10.1109/ISCAS.2018.8351786.
- [27] A. Graves. “Generating sequences with recurrent neural networks”. In: *arXiv preprint arXiv:1308.0850* (2013).
- [28] J. Kim and S. Lee. “Deep learning of human visual sensitivity in image quality assessment framework”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017, pp. 1676–1684.
- [29] A. Mittal, A. K. Moorthy, and A. C. Bovik. “No-reference image quality assessment in the spatial domain”. In: *IEEE Transactions on image processing* 21.12 (2012), pp. 4695–4708.
- [30] X. Min et al. “Blind image quality estimation via distortion aggravation”. In: *IEEE Transactions on Broadcasting* 64.2 (2018), pp. 508–517.
- [31] W. Zhang et al. “Blind image quality assessment using a deep bilinear convolutional neural network”. In: *IEEE Transactions on Circuits and Systems for Video Technology* 30.1 (2018), pp. 36–47.
- [32] Y. Zhou et al. “Omnidirectional Image Quality Assessment by Distortion Discrimination Assisted Multi-Stream Network”. In: *IEEE Transactions on Circuits and Systems for Video Technology* 32.4 (2022), pp. 1767–1777. DOI: 10.1109/TCSVT.2021.3081162.

# A Novel Voice Feature AVA and its Application to the Pathological Voice Detection Through Machine Learning

Abdulrehman Altaf<sup>1\*</sup>, Hairulnizam Mahdin<sup>2\*</sup>, Ruhaila Maskat<sup>3\*</sup>,  
Shazlyn Milleana Shaharudin<sup>4</sup>, Abdullah Altaf<sup>5</sup>, Awais Mahmood<sup>6</sup>  
Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn  
Malaysia, Batu Pahat, Johor, Malaysia<sup>1,2</sup>  
Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA  
Shah Alam, Selangor, Malaysia<sup>3</sup>  
Faculty of Science and Mathematics, Universiti Pendidikan,  
Sultan Idris, Tanjong Malim, 35900, Perak, Malaysia<sup>4</sup>  
Faculty of Computer Science and Information Technology,  
Universiti Tun Hussein Onn, Malaysia, Batu Pahat, Johor, Malaysia<sup>5</sup>  
College of Applied Computer Science, King Saud University, Riyadh, Saudi Arabia<sup>6</sup>

**Abstract**—Voice pathology is a universal problem which must be addressed. Traditionally, this malady is treated by using the surgical instruments in the varied healthcare settings. In the current era, machine learning experts have paid an increasing attention towards the solution of this problem by exploiting the signal processing of the voice. For this purpose, numerous voice features have been capitalized to classify the healthy and pathological voice signals. In particular, Mel-Frequency Cepstral Coefficients (MFCC) is a widely used feature in speech and audio signal processing. It denotes spectral characteristics of a voice signal, particularly of human speech. The modus operandi of MFCC is too time-consuming, which goes against the hasty and urgent nature of the modern times. This study has developed a yet another voice feature by utilizing the average value of the amplitudes (AVA) of the voice signals. Moreover, Gaussian Naive Bayes classifier has been employed to classify the given voice signals as healthy or pathological. Apart from that, the dataset has been acquired from the SVD (Saarbrücken Voice Database) to demonstrate the workability of the proposed voice feature and its usage in the classifier. The machine experimentation rendered very promising results. Particularly, Recall, F1 and accuracy scores obtained, are 100%, 83% and 80%, respectively. These results vividly imply that the proposed classifier can be installed in various healthcare settings.

**Keywords**—Pathological voice; healthy voice; voice feature; amplitudes; machine learning

## I. INTRODUCTION

People whose professions cause them to speak louder than normal, often suffer from some kind of voice pathology. These people may include lawyers, auctioneers, motivational speakers, legislators, singers, teachers, etc. This pathology, in turn, leads to tiredness, infections of voice tissue, face soreness, muscular dystrophy and others [1]. Apart from that, this pathology casts a negative impact upon the voice functionality and vibration regularity which sometimes leads to the increment in the vocal noise. Normal voices turned to be weak, tense, and hoarse which influences quality of voice [2]. Traditionally, voice pathology detection methods are

tendentious in their character and orientation. They are based on subjective matters [3]. For instance, in the different hospital settings, an auditory-perceptual assessment is employed which includes visual laryngostroboscopy assessment [4]. In this painful process of diagnosis, a series of clinical examinations are employed for the auditory-perceptual parameters to appraise the severity of the voice malady [5]. These appraisals are subjective and are very sensitive to the sensitivity of the parameters involved. Moreover, they happen to be very much time consuming which is not in line with the current standards of quality [6]. One more disadvantage of this method is that the patients have to be present in the hospital physically which is, of course, not feasible for the patients with critical conditions.

In sharp contrast to that, there exists an objective evaluation of the voice pathology using signal processing of the voice. In particular, the signals of patients' voice are processed to conclude whether the patient concerned is suffering from the pathology or not? No surgical treatment is employed in this method. Moreover, this procedure can also work upon the inaudible sounds [1]. These methods do not depend upon the human decisions. A patient needs not be physically present in the healthcare centres since only his/her voice is required to reach to the decision. So, the voice recording can also be shared through the internet. Upon surveying the relevant literature, one will find that different voice pathology databases exist for the sake of objective evaluation of voice pathologies. Among these, the most common include Arabic Voice Pathology Database (AVPD) [7], Saarbrücken Voice Database (SVD) [8] and the Massachusetts Eye and Ear Infirmary Database (MEEI) [9]. Upon surveying the literature about the pathological voice detection, researchers have used varied voice features and diverse machine learning classifiers for discriminating between the pathological and healthy voice signals [10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21]. The work [10] wrote a robust voice pathology detection algorithm using the theory of Deep Learning. In order to maximize the accuracy of classification, the pre-trained Convolutional Neural Network

(CNN) has been employed over the dataset of voice pathology. Besides, this work used SVD as a dataset for testing their work. The accuracy claimed by the reported study is 95.41%. Moreover, F1-Score and Recall scores were calculated to be 94.22% and 96.13%, respectively. In an other study [11], the voice pathology detection system was developed in the realm of smart healthcare. In this particular work, the voice data was taken through the IoT gadgetry, i.e., electroglottography (EGG) and microphones for capturing the EGG and voice signals. The voice feature spectrogram was employed in this particular study. These spectrograms were got from the reported signals and were given as an input to the pre-trained CNN. Moreover, the features obtained through the usage of CNN were mixed and later on processed through short long-term memory network (bi-directional in nature). The accuracy claimed by the said study was 95.65%. In an another research [12], the authors employed an Online Sequential Extreme Learning Machine (OSELM) as the classification algorithm in their work for detection of pathological voice signals. An other prominent feature of this study is the employment of long sentences instead of the single vowel letters for the sake of discrimination between the pathological and healthy voice signals. Three types of voice pathologies were addressed namely cyst, polyp, and paralysis. The accuracy achieved was 91.17%. Apart from that, precision and recall scores were 94% and 91%, respectively. The work is reported to give a high capability for detecting the pathological voice signals in the real-time clinical settings.

Many voice features have been discovered by the academicians and other researchers. Some of these include formant frequency [22]. Formant frequencies are a sort of resonance frequencies. These frequencies change with various vocal tract configurations [23]. Commonly, these formants denote the spectral contribution of the given resonances. Apart from that, peaks of these spectra about the local tract responses refer to the corresponding formants. The various plots of these formants depict the different peaks at the various frequencies. Spectrogram of a voice signals [24] is yet another voice feature. They are a kind of a waveform comprising of various events which change as the time goes by. Owing to the fact that they vary with the time, hence they fluctuate and exhibit the spectral properties. This is the reason that a single Fourier transform [25], [26] is humble to capture such kind of speedy time varying signals. Hence, for this purpose, a short-time Fourier transform (STFT) was used. STFT comprises of different Fourier transform for the pieces of the given waveform. The feature of linear predictive coding (LPC) is also used by machine learning experts to differentiate between the pathological and healthy voice signals [27], [28]. Initially, LPC was designed for compressing the digital signals for the efficient storage and transmission of the digital data. In current times, this feature is frequently being employed to draw a line of discrimination between the healthy and pathological voice signals. Moreover, this method models vocal tract in the form of linear all-pole infinite impulse response (IIR) filter.

Calculation of these features is mathematically intensive. Apart from that, they consume a lot of precious processing time which is not in line with the demands of the current era. So, we require simple but powerful voice features to do the job.

In this work, a novel voice feature by observing the behavior of amplitudes of the voice signals has been discovered. In particular, the average value of the amplitudes of the voice signal has been determined. This average value is potent enough to differentiate between the healthy and pathological voice signals. Moreover, this feature has been embedded in the machine learning algorithm to classify the two kinds of signals. The machine experimentation rendered very competitive results. Moreover, these results are better than many of those published in the literature.

Having said that, the following salient features characterize the contribution of this work to the exciting field of voice signal processing and machine learning:

- A novel voice feature based on the average value of amplitudes of the given voice signals has been determined.
- The voice feature found in the above bullet has been exploited in the machine learning algorithm to draw a rather clearer line of demarcation between the pathological and healthy voice signals.
- The proposed method rendered very competitive results. Moreover, it beats many results of the published works.

Rest of the paper has been fashioned like this. In Section II, the particular modus operandi employed in MFCC feature extraction has been explained. Section III describes the proposed methodology. In particular, the way novel voice feature has been determined, has been explained in detail. Afterwards, the reported feature has been embedded in the proposed framework to differentiate between the healthy and pathological voice signals. In Section IV, the results have been described and compared with the other published researches in the literature. Section V closes the paper with the concluding remarks and other possible research directions.

## II. RELATED WORK

Traditionally, MFCC has been employed by the machine learning experts to draw a line of demarcation between the healthy and pathological voice signals. MFCC is actually a feature selection method which plays a very critical role to distinguish the pathological voice signals from their healthy counterparts. Normally, three kinds of features are there for the recognition of the sound patterns. They are time domain, frequency domain and time-frequency domain [29]. Cepstral domain features are retrieved after taking their fast Fourier transform (FFT) of the amplitude's logarithm from spectrum data [30]. Since MFCCs closely resemble human auditory system, so their inherent power is normally harnessed for the speech recognition in the diverse problems [31]. MFCCs are normally got through power spectrum of sound signals with the short-term windowing after taking cosine transform of logarithmic power spectrum over Mel filter banks [32]. A standard modus operandi for extracting the MFCCs from the given audio signals have been depicted in the Fig. 1. Firstly, windowing functions, like Hanning and Hamming windows, are normally employed through some degree of overlapping for capturing local spectral characteristics. Secondly, various signals from the different frames are subjected to operation of

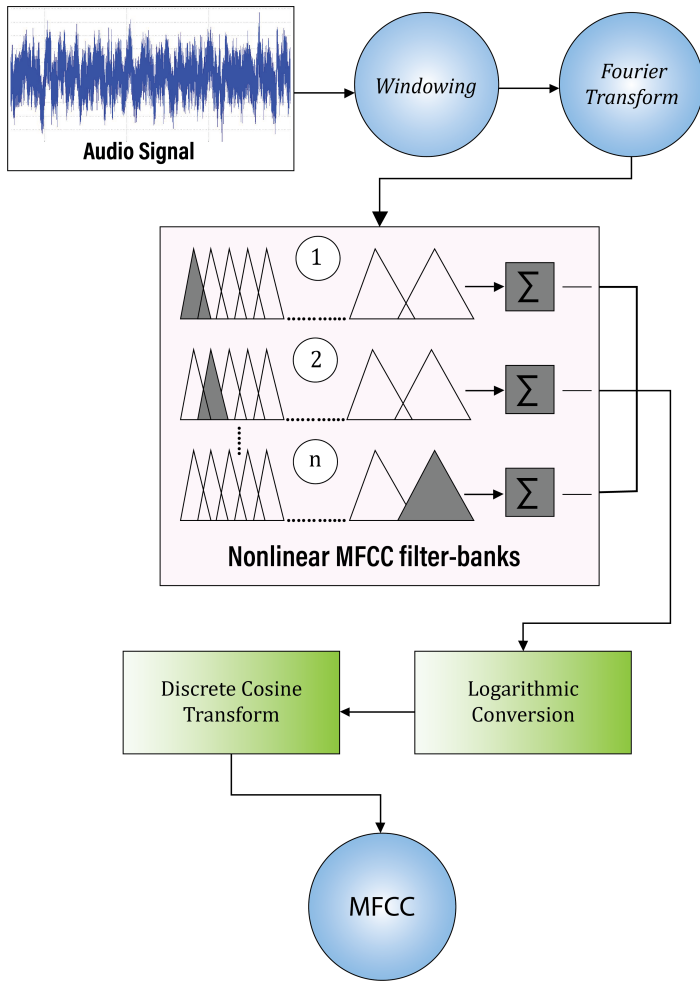


Fig. 1. MFCC feature extraction mechanism, given audio signals have been bifurcated into the overlapping frames along with some fixed intervals and weights have been given through hamming window.

discrete Fourier transform (DFT) for the sake of extracting information from the frequency domain. Thirdly, frequency domain data are filtered through a number of bandpass filters which are normally equal to designed number of the MFCC resolution (64,128,256,512). It is to be noted that centre frequencies of bandpass filters are spaced in uniformity on Mel scale  $M(f)$  [33].

$$M(f) = \frac{1000 \ln(1 + \frac{f}{700})}{\ln(1 + \frac{1000}{700})} \approx 1127 \ln(a + \frac{f}{700}) \quad (1)$$

In this equation,  $f$  denotes frequency term and  $M(f)$  denotes Mel scale. Moreover, this equation converts boundaries of filter bank to Mel scale. As soon as centre frequencies are distributed in a uniform fashion on Mel scale, values are converted back to frequency domain which renders the triangular filters. After that energies  $MF(t)$  of corresponding filter banks are computed by taking sum of energies in bandpass filters. Finally, MFCC coefficients are found through the application of discrete cosine transform (DCT) to the filtered energies from

the triangular bandpass filters [34].

$$MFCC_{i,j} = \frac{1}{T} \sum_{k=1}^T \log[MF(k)] \cos[\frac{2\pi}{T} (k + \frac{1}{2})j] \quad (2)$$

In this equation,  $MFCC_{i,j}$  refers to the  $j^{th}$  MFCC coefficient of  $i^{th}$  frame. Apart from that,  $1 \leq i \leq N$  and  $1 \leq j \leq M$ . They represent the indices of MFCC. Moreover,  $MF(k)$  is Mel filter bank amplitude of  $k^{th}$  filter. Apart from that, Table I sheds light on the varied studies carried out. In this table, one can examine the different studies based on the number of samples taken, phonemes, pathological condition of the patients, the classifier employed, the feature used and lastly the findings. Here will describe few studies in more details. In study [35], normal and pathological samples taken were 60 and 402, respectively. Besides, the vowels were taken as phonemes to apply the classifier. Apart from that, the voices of the patients were suffering from the pathological conditions of structure lesions and neoplasm. Additionally, the classifiers selected for this particular study were Support Vector Machine (SVM), Gaussian Mixture Modelling (GMM) and Deep Neural Network (DNN). The feature upon which the distinction was made between the healthy and pathological voice was MFCC. As far as the findings and outcomes of this study were concerned, SVM outperformed GMM. Besides, the classifier DNN rendered the highest accuracy. The study [36] took 56 normal samples of voices and 67 pathological samples. Moreover, the phonemes employed in this particular study were 'ah'. The pathological conditions of the patients concerned were that they were suffering from the Parkinson's disease, Vocal cord paralysis and cerebral demyelination. The classifier and the feature selected were SVM and MFCC. The accuracy obtained in this study was 93%. The last row of this table describes these parameters for the proposed study. It is to be noted that, we employed AVA as a voice feature for the sake of classification between the healthy and the pathological voice signals.

One can note that this traditional method of extracting the voice feature is very complicated and mathematically intensive. We require simple but powerful voice features to draw a clearer line of demarcation between the given voice signals.

### III. PROPOSED METHODOLOGY

In this work, a novel voice feature consisting of average value of the amplitudes (AVA) of the given voice signal has been determined. This voice feature has been, in turn, employed to distinguish between the given healthy and pathological voice signals. Fig. 2 draws the amplitudes of the healthy and pathological voice signals. Fig. 2a and 2b refer to the signals for the healthy and pathological voices. One can clearly observe that amplitudes in the positive and negative sides for the healthy signals are greater than their pathological counterparts. This is the very observation through which the novel voice feature has been determined. In particular, the average value of the amplitudes (both positive and negative) of the healthy and pathological voice signals has been found. Their average values have been further averaged. This average has been used while training our model. Once the model gets trained, testing phase have been employed.

TABLE I. OVERVIEW TABLE

Sr. #	Study	Samples	Phonemes	Pathological condition	Classifier	Feature	Findings
1	Ref. [35]	Normal: 60 Pathological: 402	Vowels	Structural lesions, neoplasm	SVM, GMM, DNN	MFCC	SVM outperforms GMM DNN provides the highest accuracy
2	Ref. [36]	Normal: 56 Pathological: 67	Vowel 'ah/	Parkinson's disease, Vocal cord paralysis cerebral demyelination	SVM	MFCC	Highest accuracy of 93%
3	Ref. [37]	Pathological: 60	Japanese vowel	Breathiness, Roughness, asthma and strain	Higher-Order Local Autocorrelation (HLAC)	Auto Regressive, (AR)-HMM, Feed Forward Neural Networks (FFNN)	87.75% accuracy
4	Ref. [38]	Normal: 53 Pathological: 602	Vowel 'ah/', Rainbow passage (German, Japanese and English)	Hyper function, Paralysis, Anterior-poster squeezing, Gastric reflux	PRAAT	Pitch, Jitter Shimmer and HNR	Efficient for English, not efficient for German and Japanese
5	Ref. [39]	Pathology: 65 Normal: 13	Spanish vowel	Dysphonia, Hyernasality and Dysarthria	Hidden Markov Model (HMM)	Nonlinear parameter, entropy	99% accuracy
6	Ref. [40]	Normal: 49 Pathological: 87	Vowel 'a/	Dysphonia	Pitch Detection Algorithm(PDA)	Pitch	Better than PRAAT
7	Proposed	Normal: 50 Pathological: 50	Vowels	Dysphonia	GaussianNB	AVA	Accuracy is 80%

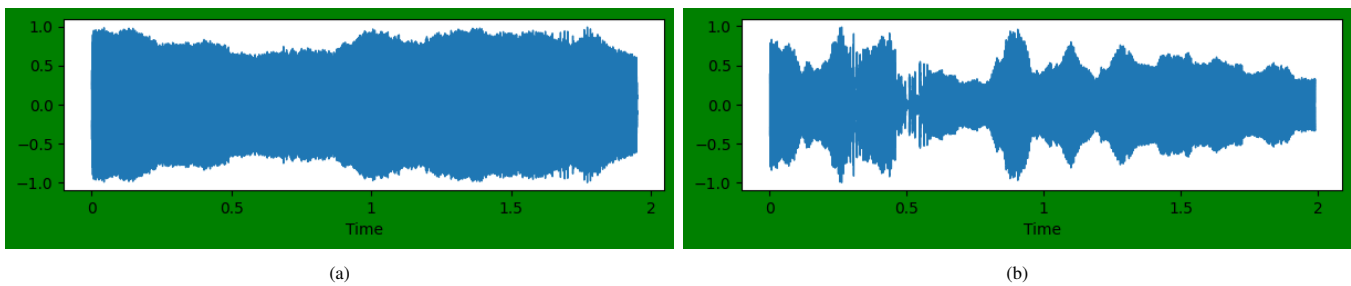


Fig. 2. Amplitudes of healthy and pathological voices: (a) Healthy voice; (b) Pathological voice.

### A. Voice Feature Based on Average Value of the Amplitudes (AVA)

This subsection finds the average value of the amplitudes of the given voice signals. The flowchart for extracting the novel voice feature AVA has been depicted in the Fig. 3. Call the Algorithm 1 with the parameters of training data ( $TD$ ), number of healthy voice files ( $P$ ) and the number of pathological voice files ( $Q$ ). Algorithm 1 works like this. The

**Algorithm 1:** AvgValue Calculation of Average Value of the Amplitudes of Voice Signals

```

Input:  $TD, P, Q$ 
Output:  $AvgValue$ 
1 for  $i \leftarrow 1$  to  $P + Q$  do
2    $[data, sampling\_rate] \leftarrow librosa.load(TD[i])$ 
3    $sum\_of\_amplitudes \leftarrow 0$ 
4   for  $j \leftarrow 1$  to  $len(data)$  do
5     if  $data[j] < 0$  then
6        $data[j] \leftarrow -data[j]$ 
7    $temp \leftarrow sum(data)$ 
8    $temp \leftarrow \frac{temp}{len(data)}$ 
9    $sum\_of\_amplitudes \leftarrow$ 
      $sum\_of\_amplitudes + temp$ 
10  $AvgValue \leftarrow \frac{sum\_of\_amplitudes}{P+Q}$ 

```

for loop of line 1 iterates for  $P + Q$  times. In each iteration, it loads the  $i^{th}$  file of the training data  $TD$  by using the load

function of the Python module librosa. It returns the stream of amplitudes  $data$  and the sampling rate  $sampling\_rate$ . Line 3 initializes a variable  $sum\_of\_amplitudes$  to zero. Lines 4 to 6 take the absolute value of the amplitudes of the voice signal carried by the array  $data$ . Lines 7 and 8 find the average value of the amplitudes and assigns this value to the variable  $temp$ . Line 9 accumulates the average values in the variable  $sum\_of\_amplitudes$ . Lastly, line 10 finds the grand average value of all the average values of the amplitudes of the given training data and assigns this value to the variable  $AvgValue$ . Algorithm 2 has been designed to train the data based on the average value found through the Algorithm 1. Line 1 invokes the Algorithm  $AvgValue$  with the parameters  $TD, P$  and  $Q$  and assigns the result to the variable  $AV$ . Lines 7 and 8 find the average value of the amplitudes of the given  $i^{th}$  file. If this value is less than the  $AV$  (line 9), this particular file is being labelled as pathological (line 10), otherwise, it is being labelled as healthy (line 12).

### B. Healthy and Pathological Voice Classifier Based on the Average Value of the Amplitudes

Proposed pathological voice classifier has been presented in the Fig. 4. On the left half of the figure, training procedure has been illustrated in a step by step fashion. The process gets sparked with the given training data which comprises of both healthy and pathological voice files. In the next stage, the amplitudes have been extracted from the given voice signals. It is followed by the calculation of the average value of the amplitudes (both in the positive and negative directions). Based

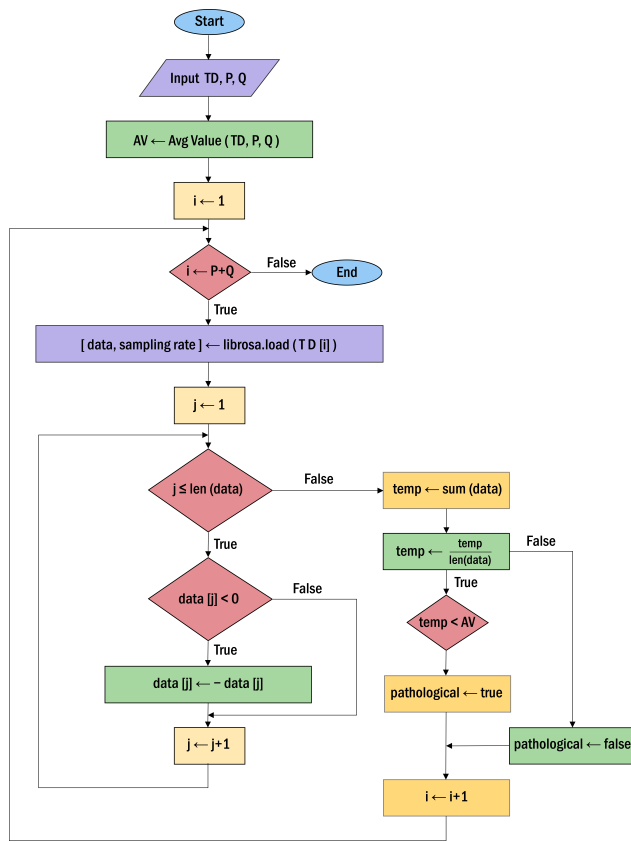


Fig. 3. Flowchart of extracting AVA.

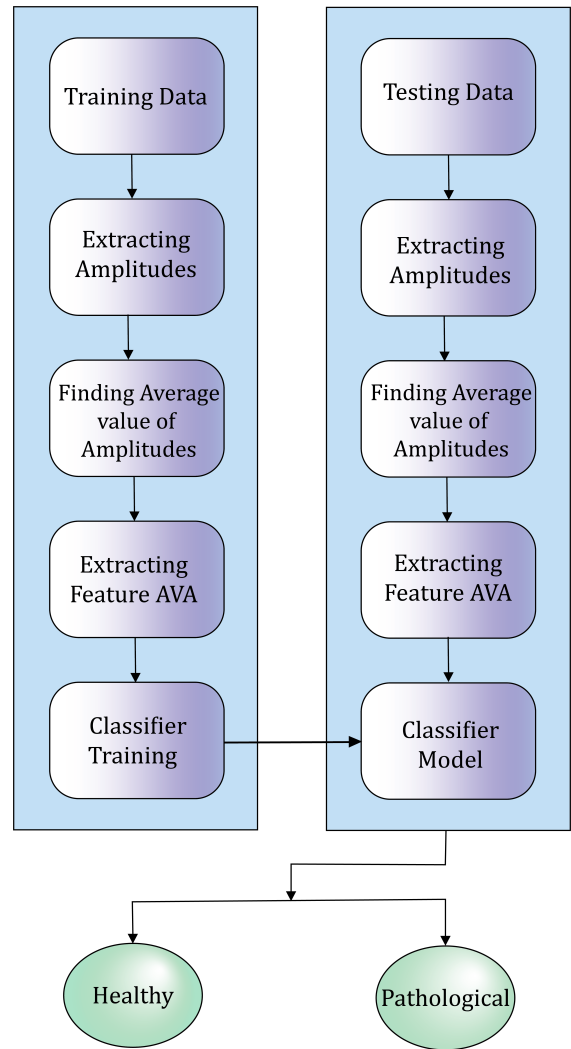


Fig. 4. AVA based pathological voice classifier.

**Algorithm 2:** Labelling voice signal as pathological or healthy based on *AvgValue*.

**Input:**  $TD, P, Q$   
**Output:** *pathological*

```

1  $AV \leftarrow AvgValue(TD, P, Q)$ 
2 for  $i \leftarrow 1$  to  $P + Q$  do
3    $[data, sampling\_rate] \leftarrow librosa.load(TD[i])$ 
4   for  $j \leftarrow 1$  to  $len(data)$  do
5     if  $data[j] < 0$  then
6        $data[j] \leftarrow -data[j]$ 
7    $temp \leftarrow sum(data)$ 
8    $temp \leftarrow \frac{temp}{len(data)}$ 
9   if  $temp < AV$  then
10     $pathological \leftarrow true$ 
11  else
12     $pathological \leftarrow false$ 

```

on this average value AVA, the GaussianNB classifier has been trained. Same process has been repeated on the right half of the figure which is pertinent to the testing data. Lastly, the classifier model outputs whether the particular voice signal is healthy or of pathological character.

#### IV. SIMULATION RESULTS

The proposed framework was simulated on the Python 3 software. We have taken 80% voice files as a training data and 20% voice files as a testing data. Additionally, these files have been taken from the SVD database. The proposed research project has used this database for the sake of experimentation. It has a collection of voice recordings of around 2000 people. To put it specifically, it contains 687 healthy voice comprising of 259 males and 428 females. Moreover, there are the recordings of 1354 pathological voices comprising of 627 males and 727 females. It is to be noted that all these recordings contain 71 different pathologies. Moreover, these recordings were sampled at 50 kHz frequency along with a 16-bit resolution. Additionally, average age of the speakers is

TABLE II. RESULTS (IN PERCENT FORM) OF DIFFERENT VALIDATION METRICS USING THE PROPOSED METHODOLOGY AND OTHER METHODS

Method	Feature	Accuracy	Precision	Recall	F1 Score	Specificity	G-mean
Ref. [10]	-	95.41	-	96.13	94.22	-	-
Ref. [11]	-	93.94	95.08	94.87	94.93	-	-
Ref. [12]	MFCC	91.17	94.0	91.0	87.0	97.67	87.55
Ref. [13]	Peak and Lag	88.70	-	88.69	-	88.71	-
Ref. [13]	Entropy	82.01	-	73.90	-	89.72	-
Ref. [14]	MPEG-7	99.994	-	73.90	-	89.72	-
Ref. [15]	LLE+CD	90.0	-	88.0	-	98.0	-
Ref. [16]	MDVP	76.0	-	45.0	-	93.0	-
Proposed	AVA	80.0	71.0	100.0	83	60.0	100.0

around 15 years. Apart from that, 1 to 3 seconds is the duration of these voice samples.

As far as the machine learning algorithm is concerned, GaussianNB algorithm was chosen to classify the healthy and pathological voice signals. Moreover, in this study, we have chosen these validation measures: accuracy, precision, recall (sensitivity), F-measures, G-mean, and specificity as shown in Eq. 3 to Eq. 7 [41], [42]. The following describes the various measures which are frequently employed in the literature.

- 1) FP (False Positive): The voice signal under consideration is of healthy character but algorithm declares it as of pathological character.
- 2) FN (False Negative): The voice signal under consideration is of pathological character but algorithm declares it as of healthy character.
- 3) TP (True Positive): The voice signal is of pathological character and algorithm declares it as of pathological character.
- 4) TN (True Negative): The voice signal is of healthy character and algorithm declares it as of healthy character.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

$$Sensitivity(Recall) = \frac{TP}{TP + FN} \quad (5)$$

$$F - Measure = \frac{2 \times Precision \times Recall}{Recall + Precision} \quad (6)$$

$$Specificity = \frac{TN}{TN + FP} \quad (7)$$

The Table II shows the results of the proposed study. Apart from that, this table also draws a comparison between the suggested work and other published researches found in the literature. As can be seen from the table, we got both the Recall and G-mean scores as 100% which validates and confirms the very idea of AVA, we conceived before launching this project. Apart from that, we got an accuracy of 80% which beats some of the published works in the literature. Moreover, one can see that various machine learning experts have used the voice feature of MFCC, Peak and Lag, Entropy, MPEG-7, LLE+CD and MDVP. Unfortunately, our study could only beat the study [16] as far as the metric of accuracy is concerned. However, our results of Recall, G-Mean and F1 score are very competitive.

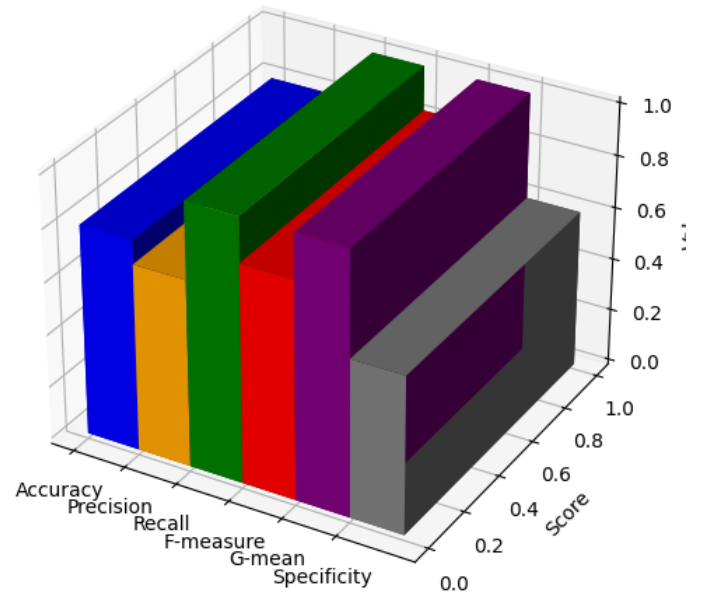


Fig. 5. Visual demonstration of validation metrics.

Besides, the Fig. 5 depicts the results of the proposed study in a graphical form which is more intuitive.

## V. DISCUSSION

Many voice features have been investigated as the literature over the pathological and healthy voice classifiers suggests. These features are input to the machine learning classifiers in order to draw a line of separation between the given pathological and healthy voice files of the different patients. The authors of this study observed a pattern after drawing and putting side by side the graphs of the pathological and healthy voices. The amplitudes of the healthy voice signals went higher as compared to their pathological counterparts. This was the very point which was further investigated. In this way, a novel voice feature termed as Average Values of the Amplitudes was found which was further imported to the machine learning classifier in order to draw a clear line of demarcation between the healthy and pathological voice signals. This study obtained Recall and G-mean scores as 100% while the accuracy achieved reached to the tune of 80%. This outcomes vividly imply that the suggested voice feature is potent enough to predict the healthy and pathological voice



signals. Moreover, we contend that the proposed voice feature can be extracted with faster speed as compared to the other features like MFCC. MFCC has twelve parameters whereas the proposed one has only one parameter. As far as the limitations of the proposed framework are concerned, it can't differentiate all the voice pathologies.

## VI. CONCLUSION

By observing an underlying pattern in the given voice signals, a novel voice feature AVA has been developed in this study based on the varying values of the amplitudes of the healthy and pathological voice signals. Although a plethora of voice features already exist when one peruses the literature but this newly developed voice feature is very simple and robust to draw a clearer line of demarcation between the healthy and the pathological voice signals. This feature has been exploited while using the machine learning algorithm to detect the pathological voices. The simulation and the machine experimentation rendered very promising results. In particular, we got the Recall and G-mean score as 100% while the accuracy achieved by this study is 80%. We assert that the proposed voice classifier can be installed in some real world healthcare setting to reap its intrinsic benefits. As a future work, other machine learning classifiers and the voice databases would be investigated to examine the workability of the novel voice feature AVA, this study has produced.

## ACKNOWLEDGMENT

I would like to express my sincere gratitude to my supervisor and the reviewers for their invaluable support and feedback. Without their guidance and input, this paper would not have been possible. Additionally, I would like to thank my family and friends for their unwavering encouragement throughout this research journey.

## REFERENCES

- [1] F. T. Al-Dhief, N. M. A. Latiff, N. N. N. A. Malik, N. S. Salim, M. M. Baki, M. A. A. Albadr, and M. A. Mohammed, "A survey of voice pathology surveillance systems based on internet of things and machine learning algorithms," *IEEE Access*, vol. 8, pp. 64514–64533, 2020.
- [2] M. A. Mohammed, K. H. Abdulkareem, S. A. Mostafa, M. Khanapi Abd Ghani, M. S. Maashi, B. Garcia-Zapirain, I. Oleagordia, H. Alhakami, and F. T. Al-Dhief, "Voice pathology detection and classification using convolutional neural network model," *Applied Sciences*, vol. 10, no. 11, p. 3723, 2020.
- [3] J. Hillenbrand, R. A. Cleveland, and R. L. Erickson, "Acoustic correlates of breathy vocal quality," *Journal of Speech, Language, and Hearing Research*, vol. 37, no. 4, pp. 769–778, 1994.
- [4] N. Saenz-Lechon, J. I. Godino-Llorente, V. Osmar-Ruiz, and P. Gómez-Vilda, "Methodological issues in the development of automatic systems for voice pathology detection," *Biomedical Signal Processing and Control*, vol. 1, no. 2, pp. 120–128, 2006.
- [5] M. Markaki and Y. Stylianou, "Using modulation spectra for voice pathology detection and classification," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2009, pp. 2514–2517.
- [6] M. S. Hossain, G. Muhammad, and A. Alamri, "Smart healthcare monitoring: a voice pathology detection paradigm for smart cities," *Multimedia Systems*, vol. 25, pp. 565–575, 2019.
- [7] N. Q. Abdulmajeed, B. Al-Khateeb, and M. A. Mohammed, "A review on voice pathology: Taxonomy, diagnosis, medical procedures and detection techniques, open challenges, limitations, and recommendations for future directions," *Journal of Intelligent Systems*, vol. 31, no. 1, pp. 855–875, 2022.
- [8] J.-N. Lee and J.-Y. Lee, "An efficient smote-based deep learning model for voice pathology detection," *Applied Sciences*, vol. 13, no. 6, p. 3571, 2023.
- [9] N. Q. Abdulmajeed, B. Al-Khateeb, and M. A. Mohammed, "A review on voice pathology: Taxonomy, diagnosis, medical procedures and detection techniques, open challenges, limitations, and recommendations for future directions," *Journal of Intelligent Systems*, vol. 31, no. 1, pp. 855–875, 2022.
- [10] M. A. Mohammed, K. H. Abdulkareem, S. A. Mostafa, M. Khanapi Abd Ghani, M. S. Maashi, B. Garcia-Zapirain, I. Oleagordia, H. Alhakami, and F. T. Al-Dhief, "Voice pathology detection and classification using convolutional neural network model," *Applied Sciences*, vol. 10, no. 11, p. 3723, 2020.
- [11] G. Muhammad and M. Alhussein, "Convergence of artificial intelligence and internet of things in smart healthcare: a case study of voice pathology detection," *Ieee Access*, vol. 9, pp. 89198–89209, 2021.
- [12] F. T. Al-Dhief, M. M. Baki, N. M. A. Latiff, N. N. N. A. Malik, N. S. Salim, M. A. A. Albader, N. M. Mahyuddin, and M. A. Mohammed, "Voice pathology detection and classification by adopting online sequential extreme learning machine," *IEEE Access*, vol. 9, pp. 77293–77306, 2021.
- [13] A. Al-Nasheri, G. Muhammad, M. Alsulaiman, Z. Ali, K. H. Malki, T. A. Mesallam, and M. F. Ibrahim, "Voice pathology detection and classification using auto-correlation and entropy features in different frequency regions," *Ieee Access*, vol. 6, pp. 6961–6974, 2017.
- [14] G. Muhammad and M. Melhem, "Pathological voice detection and binary classification using mpeg-7 audio features," *Biomedical Signal Processing and Control*, vol. 11, pp. 1–9, 2014.
- [15] J. D. Arias-Londono, J. I. Godino-Llorente, N. Sáenz-Lechón, V. Osmar-Ruiz, and G. Castellanos-Domínguez, "Automatic detection of pathological voices using complexity measures, noise parameters, and mel-cepstral coefficients," *IEEE Transactions on biomedical engineering*, vol. 58, no. 2, pp. 370–379, 2010.
- [16] A. Al-Nasheri, G. Muhammad, M. Alsulaiman, Z. Ali, T. A. Mesallam, M. Farahat, K. H. Malki, and M. A. Bencherif, "An investigation of multidimensional voice program parameters in three different databases for voice pathology detection and classification," *Journal of Voice*, vol. 31, no. 1, pp. 113–e9, 2017.
- [17] R. Ranjbarzadeh, S. Dorosti, S. Jafarzadeh Ghouschi, S. Safavi, N. Razmjoo, N. Tataei Sarshar, S. Anari, and M. Bendecheche, "Nerve optic segmentation in ct images using a deep learning model and a texture descriptor," *Complex & Intelligent Systems*, vol. 8, no. 4, pp. 3543–3557, 2022.
- [18] C. Yan and N. Razmjoo, "Kidney stone detection using an optimized deep believe network by fractional coronavirus herd immunity optimizer," *Biomedical Signal Processing and Control*, vol. 86, p. 104951, 2023.
- [19] M. Naeem, W. K. Mashwani, M. Abiad, H. Shah, Z. Khan, and M. Aamir, "Soft computing techniques for forecasting of covid-19 in pakistan," *Alexandria Engineering Journal*, vol. 63, pp. 45–56, 2023.
- [20] N. Razmjoo, V. V. Estrela, and H. J. Loschi, "Entropy-based breast cancer detection in digital mammograms using world cup optimization algorithm," in *Research Anthology on Medical Informatics in Breast and Cervical Cancer*. IGI Global, 2023, pp. 645–665.
- [21] K. Shojaei and M. Abdolmaleki, "Saturated observer-based adaptive neural network leader-following control of n tractors with n-trailers with a guaranteed performance," *International Journal of Adaptive Control and Signal Processing*, vol. 35, no. 1, pp. 15–37, 2021.
- [22] P. Singh, M. Sahidullah, and G. Saha, "Modulation spectral features for speech emotion recognition using deep neural networks," *Speech Communication*, vol. 146, pp. 53–69, 2023.
- [23] R. Islam, M. Tarique, and E. Abdel-Raheem, "A survey on signal processing based pathological voice detection techniques," *IEEE Access*, vol. 8, pp. 66749–66776, 2020.
- [24] H. Chen, L. Ran, X. Sun, and C. Cai, "Sw-wavenet: Learning representation from spectrogram and wavegram using wavenet for anomalous sound detection," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.

- [25] T. Kaneko, K. Tanaka, H. Kameoka, and S. Seki, "istftnet: Fast and lightweight mel-spectrogram vocoder incorporating inverse short-time fourier transform," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2022, pp. 6207–6211.
- [26] M. S. Khan, N. Salsabil, M. G. R. Alam, M. A. A. Dewan, and M. Z. Uddin, "Cnn-xgboost fusion-based affective state recognition using eeg spectrogram image analysis," *Scientific Reports*, vol. 12, no. 1, p. 14122, 2022.
- [27] G. Aggarwal, K. Jhajharia, J. Izhar, M. Kumar, and L. Abualigah, "A machine learning approach to classify biomedical acoustic features for baby cries," *Journal of Voice*, 2023.
- [28] M. Du, S. Liu, T. Wang, W. Zhang, Y. Ke, L. Chen, and D. Ming, "Depression recognition using a proposed speech chain model fusing speech production and perception features," *Journal of Affective Disorders*, vol. 323, pp. 299–308, 2023.
- [29] L. Jing, M. Zhao, P. Li, and X. Xu, "A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox," *Measurement*, vol. 111, pp. 1–10, 2017.
- [30] A. Abeyasinghe, M. Fard, R. Jazar, F. Zambetta, and J. Davy, "Mel frequency cepstral coefficient temporal feature integration for classifying squeak and rattle noise," *The Journal of the Acoustical Society of America*, vol. 150, no. 1, pp. 193–201, 2021.
- [31] S. Chachada and C.-C. J. Kuo, "Environmental sound recognition: A survey," *APSIPA Transactions on Signal and Information Processing*, vol. 3, p. e14, 2014.
- [32] A. Abeyasinghe, M. Fard, R. Jazar, F. Zambetta, and J. Davy, "Mel frequency cepstral coefficient temporal feature integration for classifying squeak and rattle noise," *The Journal of the Acoustical Society of America*, vol. 150, no. 1, pp. 193–201, 2021.
- [33] M. S. Hossain and G. Muhammad, "Environment classification for urban big data using deep learning," *IEEE Communications Magazine*, vol. 56, no. 11, pp. 44–50, 2018.
- [34] S. Tiwari, V. Sapra, and A. Jain, "Heartbeat sound classification using mel-frequency cepstral coefficients and deep convolutional neural network," in *Advances in Computational Techniques for Biomedical Image Analysis*. Elsevier, 2020, pp. 115–131.
- [35] S.-H. Fang, Y. Tsao, M.-J. Hsiao, J.-Y. Chen, Y.-H. Lai, F.-C. Lin, and C.-T. Wang, "Detection of pathological voice using cepstrum vectors: A deep learning approach," *Journal of Voice*, vol. 33, no. 5, pp. 634–641, 2019.
- [36] C. Vikram and K. Umarani, "Pathological voice analysis to detect neurological disorders using mfcc and svm," *Int. J. Adv. Electr. Electron. Eng.*, vol. 2, no. 4, pp. 87–91, 2013.
- [37] A. Sasou, "Automatic identification of pathological voice quality based on the grbas categorization," in *2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. IEEE, 2017, pp. 1243–1247.
- [38] S. Shinohara, Y. Omiya, M. Nakamura, N. Hagiwara, M. Higuchi, S. Mitsuyoshi, and S. Tokuno, "Multilingual evaluation of voice disability index using pitch rate," *Adv. Sci. Technol. Eng. Syst. J.*, vol. 2, no. 3, pp. 765–772, 2017.
- [39] M. Sarria-Paja and G. Castellanos-Domínguez, "Robust pathological voice detection based on component information from hmm," in *Advances in Nonlinear Speech Processing: 5th International Conference on Nonlinear Speech Processing, NOLISP 2011, Las Palmas de Gran Canaria, Spain, November 7-9, 2011. Proceedings 5*. Springer, 2011, pp. 254–261.
- [40] M. R. Jamaludin, S. H. Salleh, T. T. Swee, K. Ahmad, A. K. Ibrahim, and K. Ismail, "An improved time domain pitch detection algorithm for pathological voice," *American Journal of Applied Sciences*, vol. 9, no. 1, p. 93, 2012.
- [41] M. A. A. Albadr, S. Tiun, F. T. Al-Dhief, and M. A. Sammour, "Spoken language identification based on the enhanced self-adjusting extreme learning machine approach," *PloS one*, vol. 13, no. 4, p. e0194770, 2018.
- [42] M. A. A. Albadr, S. Tiun, M. Ayob, F. T. Al-Dhief, K. Omar, and F. A. Hamzah, "Optimised genetic algorithm-extreme learning machine approach for automatic covid-19 detection," *PloS one*, vol. 15, no. 12, p. e0242899, 2020.

# Improved Model for Smoke Detection Based on Concentration Features using YOLOv7tiny

Yuanpan ZHENG<sup>1\*</sup>, Liwei Niu<sup>2</sup>, Xinxin GAN<sup>3</sup>, Hui WANG<sup>4</sup>, Boyang XU<sup>5</sup>, Zhenyu WANG<sup>6</sup>  
Zhengzhou University of Light Industry, Zhengzhou 450000, Henan, China<sup>1,2,4,6</sup>  
SIPPR Engineering Group Co., Ltd, Zhengzhou 450007, Henan, China<sup>3</sup>  
Zhengzhou University of Industrial Technology, Zhengzhou 451100, Henan, China<sup>5</sup>

**Abstract**—Smoke is often present in the early stages of a fire. Detecting low smoke concentration and small targets during these early stages can be challenging. This paper proposes an improved smoke detection algorithm that leverages the characteristics of smoke concentration using YOLOv7tiny. The improved algorithm consists of the following components: 1) utilizing the dark channel prior theory to extract smoke concentration characteristics and using the synthesized  $\alpha$ RGB image as an input feature to enhance the features of sparse smoke; 2) designing a light-BiFPN multi-scale feature fusion structure to improve the detection performance of small target smoke; 3) using depth separable convolution to replace the original standard convolution and reduce the model parameter quantity. Experimental results on a self-made dataset show that the improved algorithm performs better in detecting sparse smoke and small target smoke, with mAP@0.5 and Recall reaching 94.03% and 95.62% respectively, and the detection FPS increasing to 118.78 frames/s. Moreover, the model parameter quantity decreases to 4.97M. The improved algorithm demonstrates superior performance in the detection of sparse and small smoke in the early stages of a fire.

**Keywords**—YOLOv7tiny; smoke detection; dark channel; smoke concentration; feature fusion; depthwise separable convolution

## I. INTRODUCTION

With the rapid development of the national economy and various industries, factories are producing more production materials, but they are also facing increased safety risks. High-density residential buildings are increasingly engaging in intensive fire and electricity usage behaviors. According to statistics from the Ministry of Emergency Management as of January 20, 2022, there were a total of 748,000 recorded fires in 2021, resulting in over 4,000 casualties and direct economic losses exceeding 6.75 billion yuan [1]. Therefore, it is crucial to research fire and smoke detection methods to ensure public property safety.

Currently, smoke detection research can be categorized into methods based on hardware sensors and wireless signals, and methods based on computer vision [2]. However, methods based on hardware sensors and wireless signals have poor adaptability in certain scenarios and do not perform well [3]. To overcome these limitations, computer vision-based smoke detection methods have been widely employed in recent years. Surveillance systems have also evolved from simulation-based, networked, and high-definition systems to intelligent systems. Now, surveillance resources are not only utilized for local monitoring functions but also integrated with computer vision for intelligent monitoring. Object detection algorithms based on deep learning have rapidly developed and become

the mainstream method for smoke detection, as they possess powerful feature learning and representation capabilities, better meeting the requirements of the big data era in comparison to traditional machine learning methods [4].

He et al. [5] proposed a deep fusion convolutional neural network for smoke detection based on efficient attention, integrating spatial and channel attention mechanisms to address the issue of detecting small smoke. Sun et al. [6] presented an improved convolutional neural network for the rapid identification of forest fire smoke. However, the algorithm has poor generalization ability and weak robustness, only exhibiting high detection capability in specific scenarios. Wang et al. [7] proposed a smoke detection algorithm based on Faster R-CNN. Firstly, smoke is extracted based on its motion features, and then the Faster R-CNN network is used to extract and recognize the smoke image features, achieving high accuracy. However, the Faster R-CNN network structure is complex, and real-time detection is poor.

In recent years, the YOLO series models have garnered extensive research in the field of object detection due to their real-time performance, one-stage detection, simplicity, and good accuracy. Ren et al. [8] implemented fire detection and recognition using an improved YOLOv3 network. The algorithm improves the accuracy and detection speed of small smoke targets by modifying the predicted box sizes of the K-means clustering algorithm in YOLOv3. Cao et al. [9] proposed a precision enhancement strategy for YOLOv4 based on multi-scale feature maps and made improvements in detecting small objects by enhancing the feature extraction network. However, this significantly increased the algorithm complexity, resulting in a significant decrease in real-time detection. Xue et al. [10] proposed an improved model based on YOLOv5s. To address the issue of capturing effective information from small-sized targets in long-distance forest fire images, transfer learning methods were used to enhance the accuracy of small-target forest fire smoke detection. However, this model has a complex structure, and the detection accuracy is not sufficient [11].

The aforementioned fire smoke detection algorithms have improved the accuracy of smoke detection to some extent. However, they still face the following difficulties in the early stages of actual fire scenarios: 1) high false negative rate for thin smoke with a slow initial spread in fires; 2) difficulty in detecting small smoke targets captured from long distances; 3) high complexity of model algorithms, making real-time detection challenging.

To address these issues, this paper proposes a YOLOv7tiny lightweight improved network based on smoke concentration features, which significantly enhances the original network for complex smoke detection scenarios. The algorithm mainly includes 1) Extracting smoke concentration features based on the atmospheric transmission principle to enhance smoke characteristics and improve the detection capability for thin smoke; 2) Using a weighted bidirectional feature fusion structure to replace the original PAN+FPN feature fusion method, enhancing the algorithm’s ability to detect small smoke targets; 3) Replacing the regular convolutions in the original network with depthwise separable convolutions with fewer parameters. The main contributions of this paper are:

1) Extracting smoke image concentration features based on the dark channel prior theory and enhancing the original RGB image to an  $\alpha$ RGB image with smoke concentration features as the network input have been proven to enhance the detection capability of early-stage fires with thin smoke through experiments.

2) Proposing a lightweight feature fusion structure (light-BiFPN) to enhance the detection of small smoke targets in the YOLOv7tiny network and reduce the false negative rate of small smoke targets.

3) Replacing the standard convolutions of the original algorithm with depthwise separable convolutions, and experimental results show a significant reduction in parameters with minimal impact on accuracy.

Finally, the superiority of the proposed improvement algorithm was confirmed by analyzing the experimental results.

4) A dataset was created for detecting smoke objects in

outdoor real-world scenes. The dataset comprises 1671 smoke images with corresponding labels indicating the position of the smoke bounding boxes. This dataset holds immense significance for researching the detection of smoke in the initial phases of actual fire scenarios.

## II. BACKGROUND

The YOLOv7 algorithm is a novel object detection algorithm introduced by the original development team of YOLOv4 in July 2022. Compared to previous versions of the YOLO series, this algorithm enhances the learning capability of the network through the use of the C5 module in the aggregation network. Additionally, it introduces attention mechanisms in the backbone feature extraction network to optimize the representation of target features, thereby achieving real-time detection. However, this algorithm has a relatively lower average precision. To achieve high-precision fire smoke detection in complex outdoor environments while reducing the number of algorithm parameters and improving detection speed, this study proposes improvements to the YOLOv7tiny algorithm. The improved algorithm includes the incorporation of a smoke concentration feature extraction structure and the use of a more lightweight multi-scale feature fusion network and optimized depthwise separable convolutions. With these enhancements, the algorithm can adapt to complex scenarios and achieve good real-time detection capability.

YOLOv7tiny is a deep learning-based object detection model composed of four parts: Input, Backbone, Neck, and Head. Fig. 1 shows the diagram of the YOLOv7tiny model. The Input part applies random mosaic data augmentation and K-means clustering to optimize the model training by designing anchor boxes for preprocessing the input images.

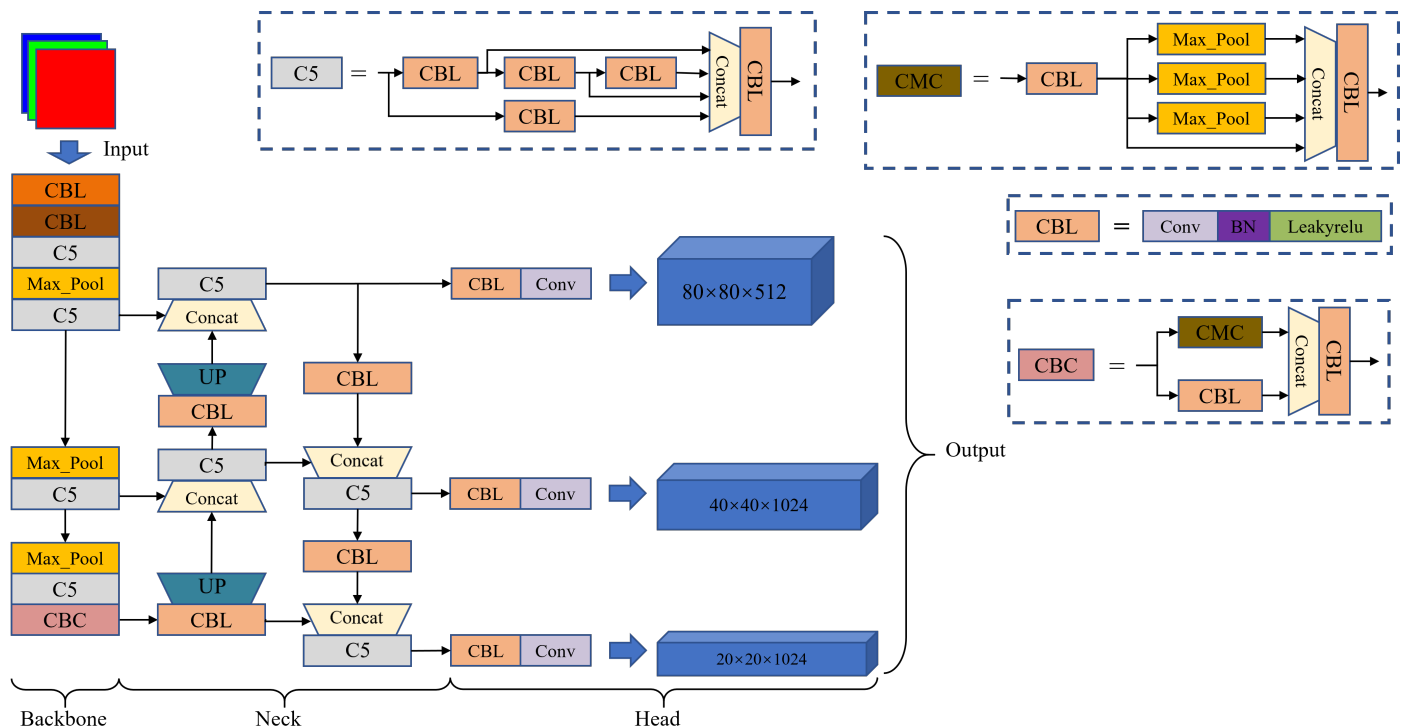


Fig. 1. The general architecture of the YOLOv7tiny network.

The Backbone part consists of multiple CBL modules, a C5 layer, and an MP layer. The CBL module is composed of a Convolution layer, a Batch Normalization layer, and a Leaky ReLU function. The C5 layer is formed by concatenating multiple CBL modules, and the MP layer includes CBL modules and Maxpool. The Neck part employs a feature fusion network, which adopts the YOLOv5 series Path Aggregation Feature Pyramid Network (PAFPN) architecture and combines Feature Pyramid Networks (FPN) [12] and Path Aggregation Networks (PAN) [13] to achieve multi-scale learning and retain small object features before downsampling. However, tensor concatenation for feature fusion lacks comprehensive integration of adjacent layer information, and nearest-neighbor interpolation for upsampling cannot effectively balance speed and accuracy in smoke detection tasks. The fusion network does not adequately focus on small object feature information, which can result in feature loss. The Head part uses a detection head similar to the YOLOR model, introducing the Implicit representation strategy [14] to refine the predictions. Based on the fused feature values, the images are classified into large, medium, and small categories, with the small image prediction branch primarily focusing on small defect objects. However, the detection head's use of IDetect to connect ordinary convolution prevents the fusion results from emphasizing the intended targets. Additionally, the detection head lacks targeted strategies to enhance small object detection performance.

### III. PROPOSED METHOD

#### A. Smoke Concentration Feature Extraction Based on Dark Channel

Smoke concentration is a characteristic of smoke that directly reflects the content of smoke in the air per unit volume. In images, smoke concentration is closely related to the transmittance of the smoke image.

$$\alpha = 1 - t \quad (1)$$

Generally, the larger the smoke concentration ( $\alpha$ ), the smaller the transmittance of the image ( $t$ ). The transmittance can be described by the smoke diffusion equation, which is a commonly used mathematical model for describing smoke concentration. Its form is as follows:

$$I = J \times t + A \times (1 - t) \quad (2)$$

Here,  $I$  represents the original foggy image,  $J$  represents the clear image after defogging,  $t$  represents the image transmission rate, and  $A$  represents the atmospheric light intensity. The dark channel prior theory [15] is a commonly used image defogging algorithm. It is based on the fog equation in Eq. (2) and analyzes the dark channel of the image to extract the transmission rate of the foggy image, thereby achieving image-defogging. The formula for the dark channel prior theory is as follows:

$$\min_{\Omega} \left( \min_C \frac{I^C}{A^C} \right) = \left\{ \min_{\Omega} \left( \min_C \frac{J^C}{A^C} \right) \right\} t + 1 - t \quad (3)$$

In the equation,  $I^C$  represents the RGB channels of the original foggy image, and  $J^C$  represents the clear and fog-free image. Through the analysis conducted by He et al. [15], it has been

revealed that the majority of images in real outdoor fog-free scenes have a significant amount of dark channels with very low pixel values, i.e.,  $\min_{\Omega} \left( \min_C \frac{J^C}{A^C} \right) \rightarrow 0$ . Therefore, after simplifying Equation (3), we can proceed with the processing:

$$t = 1 - \min_{\Omega} \left( \min_C \frac{I^C}{A^C} \right) \quad (4)$$

In the equation,  $\Omega$  represents the sliding window size. First, the brightest region is searched in the dark channel image, and then the brightness of the corresponding region in the original image is taken as the atmospheric light intensity ( $A^C$ ). As a result, the transmission rate of the foggy image ( $t$ ) can be calculated.

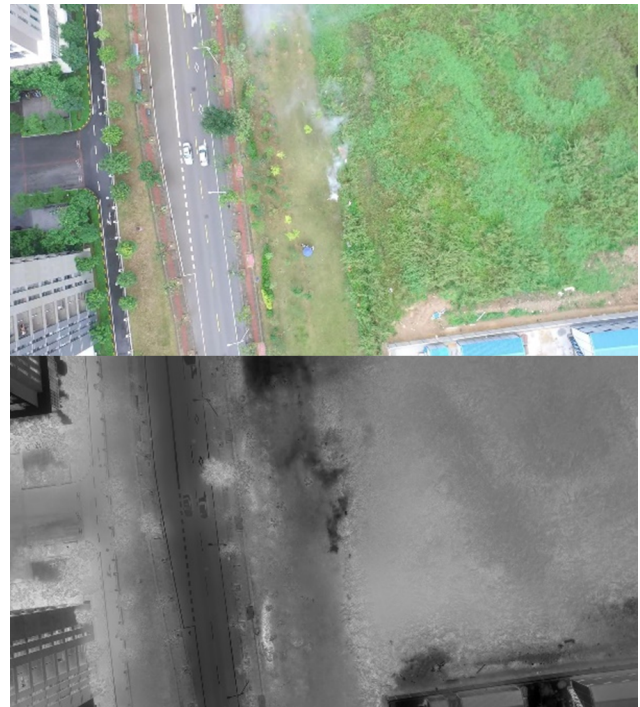


Fig. 2. Smoke image and corresponding transmittance grayscale image.

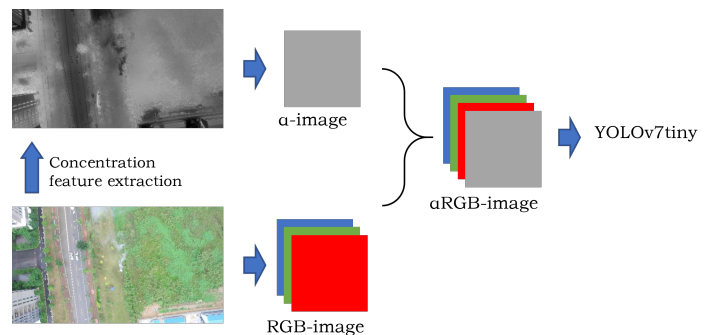


Fig. 3. Extraction of concentration features.

Mo et al. [16] extracted smoke transmittance based on the smoke aerosol equation and measured smoke concentration under different lighting conditions. The experiments demonstrated that using smoke transmittance for estimating smoke

concentration is feasible and accurate. According to Eq. (4), the transmittance of smoke images can be calculated pixel by pixel. Mapping the transmittance of smoke to a grayscale image allows for a visual representation of the transmittance map as shown in Fig. 2 (bottom).

The transmittance grayscale image exhibits dark areas that indicate low transmittance, suggesting a blockage of light, similar to smoke particles. Consequently, the smoke transmittance image is combined with the RGB image of the smoke, creating a four-dimensional vector as the input for the network model. This merged  $\alpha$ RGB image, as depicted in Fig. 3, retains the original image's shape, color, and texture, while also reflecting the inherent concentration features of the smoke.

### B. Improved Feature Fusion Network

Fig. 4 shows the PAN, BiFPN, and light-BiFPN feature fusion structures. In object detection tasks, feature fusion plays a crucial role in enhancing model accuracy. Traditional feature fusion methods focus on top-down and bottom-up feature propagation processes, with the PAN structure (Fig. 4(a)) being the most representative method [13]. By cascading, the PAN structure merges feature information from different levels and scales to expand the model's receptive field and improve detection accuracy. Its main advantage lies in effectively leveraging information from features of various scales to obtain a richer and more accurate representation.

However, the PAN structure does have some deficiencies when dealing with small objects, which can be manifested in the following two aspects:

- 1) **Feature Information Loss:** When merging feature information from different levels and scales, the PAN structure is prone to information loss, especially impacting the detection performance of small objects.
- 2) **Unstable Fusion Effects:** The cascading approach utilized in the PAN structure tends to encounter problems like gradient vanishing or explosion, leading to unstable feature fusion effects.

To address these issues, this study replaces the original PAN structure with the light-BiFPN (Bidirectional Feature Pyramid Network) [17]. The BiFPN structure [Fig. 4(b)] introduces lateral connections during the top-down and bottom-up fusion processes, effectively enhancing the exchange and transmission of feature information, particularly improving the detection performance of small targets.

The BiFPN structure is composed of multiple cascaded BiFPN modules. Each module comprises two feature propagation paths (top-down and bottom-up) and lateral connection paths. During the feature propagation process, the BiFPN module adopts a multi-level feature fusion approach to combine features from multiple sizes. These fused features are then passed to the subsequent module until the final module outputs the ultimate feature map. The lateral connection paths employ learnable weights to facilitate effective feature fusion between different layers. The weights of lateral connections are obtained through convolutional operations. Assuming the input feature map is  $x_i$ , the weights of lateral connections  $w_{ij}$  can be denoted as:

$$w_{ij} = ReLU(W_{ij}[x_i, x_j]) \quad (5)$$

Here,  $W_{ij}$  represents a learnable weight matrix, and  $[x_i, x_j]$  signifies the concatenation of feature maps  $x_i$  and  $x_j$ .

Compared to the PAN structure, the BiFPN structure better preserves detailed information of small targets, thereby improving the accuracy and robustness of object detection. Additionally, due to the scalability of the BiFPN structure, different numbers of modules and structural parameters can be chosen according to the specific scenario to achieve optimal detection performance. The light-BiFPN used in this study [Fig. 4(c)] reduces the feature layers P6 and P7 to reduce model parameters and optimize model speed.

### C. Depthwise Separable Convolution

In object detection algorithms, convolutional neural networks (CNN) are commonly employed as backbone networks to extract image features. Standard convolution serves as

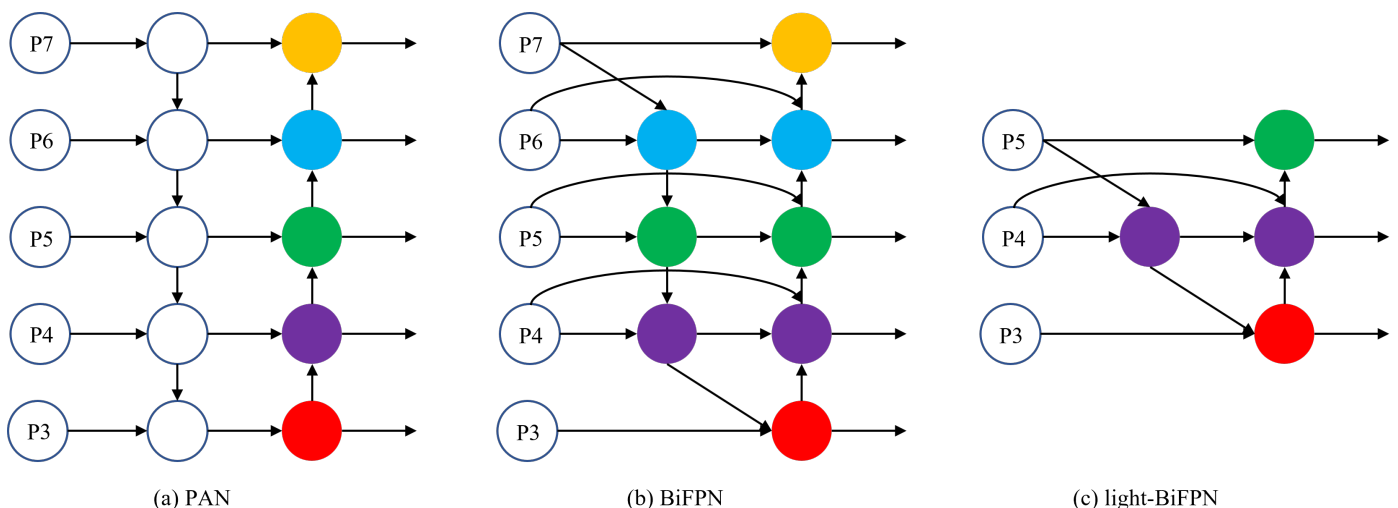


Fig. 4. PAN, BiFPN, and light-BiFPN feature fusion structures.

one of the most prevalent CNN modules, extracting features by conducting convolution operations on input feature maps and convolution kernels. However, when dealing with large-scale feature maps, standard convolution leads to significantly increased computational and memory consumption, restricting the depth and complexity of the model.

To address this issue, this paper adopts Depthwise Separable Convolution (DSC) [18] as a replacement for standard convolution in the backbone network. Depthwise Separable Convolution decomposes the standard convolution into depthwise convolution and pointwise convolution, performing convolution operations on each channel and each pixel of the input feature map, respectively. This approach considerably reduces the number of parameters and computations while ensuring the accuracy and efficiency of the model.

Depthwise Separable Convolution can be represented by the following formula:

$$Y = PW(DW(X)) \quad (6)$$

Where  $X$  represents the input feature map,  $DW$  represents the depth convolution operation,  $PW$  represents the pointwise convolution operation, and  $W$  represents the parameters of the convolutional kernel. The depth convolution operation and the pointwise convolution operation correspond to two independent convolutional layers, with parameter quantities of  $D_k$  and  $D_k \times D_o$ , respectively. Here,  $D_k$  represents the number of channels in the input feature map,  $K$  represents the size of the convolutional kernel, and  $D_o$  represents the number of channels in the output feature map. Compared to standard convolution, depthwise separable convolution reduces the parameter and computational requirements by  $K^2$  and  $D_k$  times, respectively.

#### IV. EXPERIMENTAL AND RESULT ANALYSIS

To test the effectiveness of the improved algorithm, training and testing were conducted on a self-made dataset. The optimization effects of various improvements were analyzed through horizontal comparative experiments and vertical ablation experiments.



Fig. 5. Sample images from a self-made fire smoke detection dataset.



Fig. 6. Example of image annotation.

### A. Dataset and Preprocessing

Currently, there is a limited availability of publicly accessible outdoor real fire smoke datasets. In this study, 3604 unlabeled smoke images were collected from publicly available smoke image datasets, as shown in Fig. 5. After removing low-quality images, 1671 smoke images were selected and manually annotated using the Labellmg tool to create a self-made smoke detection dataset in Pascal VOC2007 format. Fig. 6 demonstrates the smoke targets and their corresponding XML information. The dataset was split as follows:

$$(TrainingSet + ValidationSet) : TestSet = 9 : 1$$

$$TrainingSet : ValidationSet = 9 : 1$$

The training set, validation set, and test set consist of 1352, 151, and 168 images, respectively.

In the data preprocessing stage of this experiment, in addition to using traditional image processing techniques such as image flipping and HSV color space enhancement, random mosaic, and mixup image processing techniques [19] were also applied to randomly augment the dataset, aiming to enhance the robustness of the model.

The random mosaic technique combines multiple images into a new image to enhance the diversity of the dataset, while the mixup technique linearly blends two different images to generate a new image. Both data augmentation techniques effectively increase the sample size of the dataset, improving the model's generalization ability and further enhancing the accuracy of smoke object detection.

### B. Experimental Environment and Parameter Settings

#### 1) Hardware and software Environment

The experimental hardware environment of this article is shown in Table I.

TABLE I. EXPERIMENTAL ENVIRONMENT

CPU	AMD EPYC 7773X @ 3.50GHz
GPU	GeForce RTX 3090
RAM	30G
Operating System	Ubuntu
Programming Language	Python 3.8
Deep Learning Framework	PyTorch 1.8
GPU Acceleration Library	CUDA 11.1

#### 2) Training Hyperparameters Settings

The experimental hyperparameter settings of this article are shown in Table II.

TABLE II. TRAINING HYPERPARAMETERS SETTINGS

Hyperparameter	Value
Mosaic Probability	0.5
Mixup Probability	0.5
Maximum Learning Rate	0.01
Minimum Learning Rate	0.0001
Epoch	300

During the training process, the VOC pre-trained weights of YOLOv7tiny were utilized. The first 50 epochs comprised

of frozen training, where only the Neck and Head parts' parameters were trained while the backbone feature extraction network remained frozen. From epoch 51 to 300, the unfrozen training stage occurred, and the entire network was trained. The batch size was set to 64 during the frozen training stage, and it was reduced to 32 during the unfrozen training stage to accommodate the increase in training parameters. The cosine learning rate decay method was employed to progressively decrease the learning rate from 0.01 to 0.0001. The stochastic gradient descent method with a momentum of 0.937 was chosen as the parameter optimizer. Additionally, a weight decay coefficient of  $5e-4$  was implemented to prevent overfitting during the training process.

### C. Evaluation Metrics

To evaluate the performance of the improved algorithm, this study uses four metrics for algorithm assessment: Recall, mean Average Precision (mAP), Frames Per Second (FPS), and model parameter count (Params).

1) *Recall*: Recall measures the detection rate of a model for all true positive samples. In smoke object detection tasks, the calculation formula for Recall is as follows:

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

Where  $TP$  represents true positive, referring to the number of positive samples correctly detected by the model, while  $FN$  represents false negative, indicating the number of positive samples that the model fails to detect. Recall is utilized in this paper as one of the evaluation metrics to assess the detection capability of the algorithm.

2) *mAP*: mAP stands for mean Average Precision, which measures the average precision of a model at different confidence thresholds. In smoke object detection tasks, the formula to calculate mAP is as follows:

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i \quad (8)$$

Where  $n$  represents the number of classes, and  $AP_i$  represents the average precision of the  $i$ th class. In this paper, since only smoke is involved as the target, mAP can be considered as AP, used to evaluate the detection accuracy of the algorithm.

3) *FPS*: FPS stands for Frames Per Second, which measures the number of frames processed by a model per unit of time. In the smoke detection task, the calculation formula for FPS is as follows:

$$FPS = \frac{1}{t} \quad (9)$$

In this case,  $t$  represents the average time for processing a frame image. This paper utilizes Frames Per Second (FPS) as one of the evaluation metrics to assess the detection speed of the algorithm.

4) *Params*: The model parameter count refers to the number of trainable parameters in the model, which is an important indicator for evaluating model complexity. In the task of smoke object detection, the calculation formula for model parameter count is given by Eq. 10.

$$N = \sum_{i=1}^n (w_i h_i c_i k_i^2 + b_i) \quad (10)$$



TABLE III. ABLATION EXPERIMENT RESULTS

Experimental Number	Improvement			Evaluation Metric			
	$\alpha$ RGB	Light-BiFPN	DSC	Recall	mAP@0.5	FPS	Params(M)
1	✗	✗	✗	88.21%	89.48%	106.65	6.23
2	✓	✗	✗	91.12%	92.33%	95.40	6.23
3	✗	✓	✗	91.99%	91.54%	94.46	6.31
4	✗	✗	✓	86.63%	87.80%	<b>124.68</b>	<b>4.82</b>
5	✓	✓	✓	<b>95.62%</b>	<b>94.03%</b>	118.78	4.97

Here,  $n$  represents the number of layers in the model.  $w_i$ ,  $h_i$ , and  $c_i$  represent the width, height, and number of channels of layer  $i$ , respectively.  $k_i$  represents the size of the convolutional kernel in layer  $i$ , and  $b_i$  represents the bias term in layer  $i$ . In this study, the model complexity is evaluated based on the number of model parameters.

#### D. Ablation Experiment

To validate the benefits of each improvement point on the network model, five ablation experiments were conducted. The experimental environment and parameter settings were kept consistent. The results of the ablation experiments are shown in Table III.

1) The first set of experiments is conducted using the YOLOv7tiny algorithm, serving as a comparative benchmark for the subsequent improvement experiments.

2) The second group of experiments is a control experiment with the inclusion of smoke concentration features. By introducing smoke concentration features, the computational burden of the model increases, resulting in a decrease in the detection frame rate. However, it achieved good performance in terms of Recall and mAP, with improvements of 2.91 and 2.85 percentage points, respectively.

3) By analyzing the experimental data of the first and third groups, it is concluded that the light-BiFPN structure increases the number of model parameters due to the addition of skip connections, which leads to a decrease in the detection frame rate. However, it demonstrates good performance in terms of accuracy and recall rate, with improvements of 3.78 and 2.06 percentage points, respectively.

4) The fourth set of experiments replaced the standard convolution in the original YOLOv7tiny network model with depthwise separable convolution (DSC). Analyzing the experimental data compared to the baseline network reveals that DSC can significantly reduce the number of parameters and improve the detection frame rate. However, the reduced number of parameters limits the expressive power of the model.

5) In the fifth experiment, the improved YOLOv7tiny network based on smoke concentration features proposed in this paper is evaluated. From the experimental data, it can be observed that compared to the baseline network, the Recall and mAP have improved by 7.41 and 4.55 percentage points, respectively. The FPS has improved by 12.13 frames/s, and the number of parameters has decreased from 6.23M to 4.97M. Therefore, it can be concluded that the algorithm proposed in this paper is lighter and more accurate.

#### E. Comparative Experiment

To investigate the performance of the improved network in detecting different targets, this study conducted three sets of comparative experiments on a self-made dataset: comprehensive comparison experiment, smoke concentration comparison experiment, and multi-scale target comparison experiment. To comprehensively assess the performance of the algorithm, mainstream object detection models were selected as the comparison models, including RetinaNet [20], CenterNet [21], EfficientDet [22], Faster R-CNN [23], SSD [24], and YOLOv5s [25].

1) *Comprehensive comparative experiment:* A comprehensive comparative experiment was conducted by training six mainstream detection algorithms on a self-made dataset for 300 epochs as comparison algorithms to the proposed algorithm in this paper. From the variations of mAP@0.5 of each algorithm during the training process (shown in Fig. 7), it can be observed that, apart from the proposed algorithm, Faster R-CNN and YOLOv5s performed remarkably well on this dataset. Both the proposed algorithm and Faster R-CNN converged quickly (basically converged at 50 epochs). The proposed algorithm achieved an mAP of 94.03% after final convergence, surpassing other algorithms.

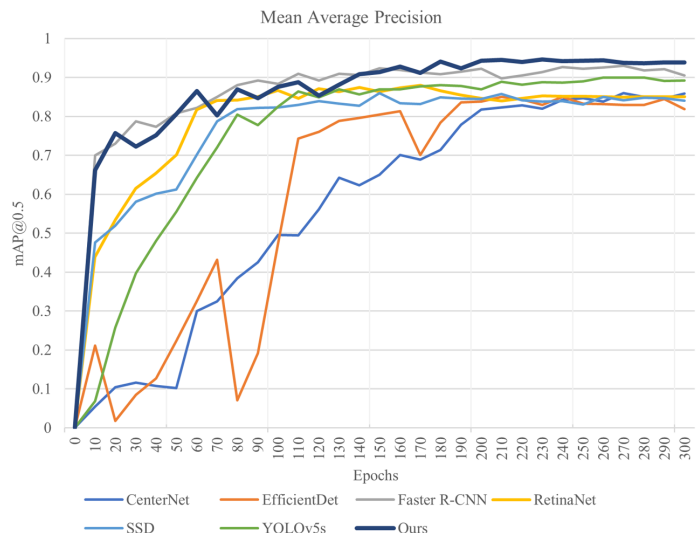


Fig. 7. mAP@0.5 Variation Graph of Each Algorithm during Training Process.

2) *Smoke Concentration Comparison Experiment:* The concentration features are extracted from the smoke images in the dataset. Then, the mean concentration of the smoke region can be obtained by calculating the average of the con-

centrations within the smoke bounding box. The distribution of smoke concentrations in this dataset is shown in Fig. 8.

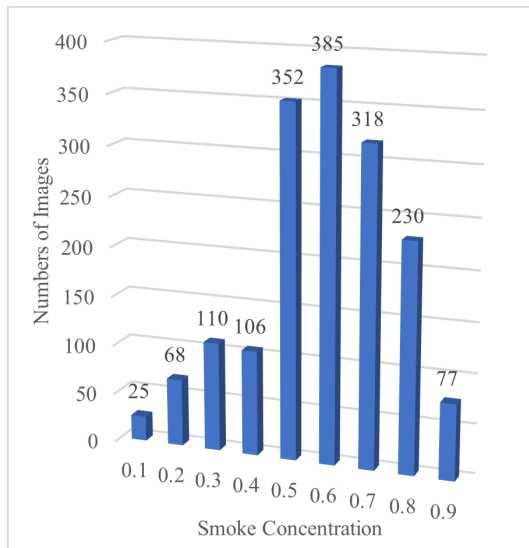


Fig. 8. Histogram of Smoke Target Concentration Distribution.

In Fig. 8, the x-axis represents the smoke concentration, and the y-axis represents the number of smoke images corresponding to each concentration. In this experiment, the smoke concentration is divided into low concentration and high concentration. The low concentration is defined as below 0.5, and the high concentration is defined as 0.5 and above. The performance of the improved model and mainstream object detection models are compared on the low-concentration and high-concentration test image sets.

TABLE IV. EXPERIMENTAL RESULTS OF SMOKE CONCENTRATION COMPARISON

Compare Models	Low Concentration mAP(%)	High Concentration mAP(%)	Params (M)
RetinaNet	84.75%	88.01%	36.33
CenterNet	83.35%	87.64%	32.67
EfficientDet	82.56%	86.51%	<b>3.83</b>
Faster R-CNN	90.45%	93.67%	136.69
SSD	82.96%	87.26%	23.61
YOLOv5s	86.75%	90.43%	46.63
Ours	<b>93.27%</b>	<b>94.64%</b>	4.97

From Table IV, it can be observed that both the mainstream algorithm models and our proposed improved algorithm model achieve similar detection accuracy for high-concentration smoke. However, our improved algorithm model achieves a significant reduction in parameter size, down to 4.97M. This reduction is particularly important for deploying the model on edge devices. Ordinary algorithms struggle to distinguish low-concentration smoke due to its semi-transparent nature. In contrast, our improved algorithm achieves good performance on low-concentration smoke, thanks to the introduced  $\alpha$ RGB concentration feature.

3) *Multi-scale Object Comparison Experiment*: In the early stage of a fire, the smoke volume is usually small. However, the detection of smoke in the early stage is particularly important for firefighting. Therefore, a small object comparison

experiment is designed to test the performance of different algorithms in detecting smoke from small objects.

Using the K-means algorithm, a cluster analysis of the size of smoke targets in the dataset was performed. The average silhouette coefficient was found to be 1.74, and the center points corresponded to large, medium, and small targets with sizes of 33×23, 80×60, and 160×142, respectively. Fig. 9 shows the distribution of width and height for the three scales of smoke targets in the self-made dataset. The horizontal axis represents the width of the smoke target, and the vertical axis represents the height of the smoke target.

The scale analysis of 168 smoke images in the test set reveals that there are 36 large objects, 47 medium-sized objects, and 85 small objects. As shown in Fig. 10, it can be observed that small smoke objects occupy a significant portion. The detection results of various algorithm models on this test set are shown in Fig. 10.

From the multi-scale object comparison experimental results in Fig. 11, it can be seen that although CenterNet, EfficientDet, SSD, and YOLOv5s have higher mAP in detecting large and medium objects, they are slightly inferior in detecting small objects. RetinaNet and Faster R-CNN perform well in detecting objects of different scales, but overall, the mAP is relatively low. By using improved algorithms, especially the optimization of light-BiFPN, the detection accuracy of small objects is significantly improved, and they have higher mAP in object detection at various scales.

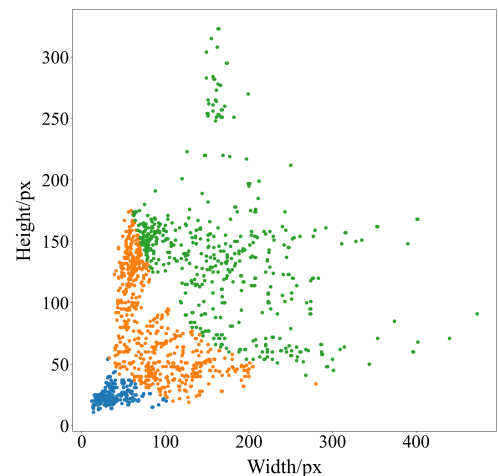


Fig. 9. Distribution of smoke objects in self-made dataset.

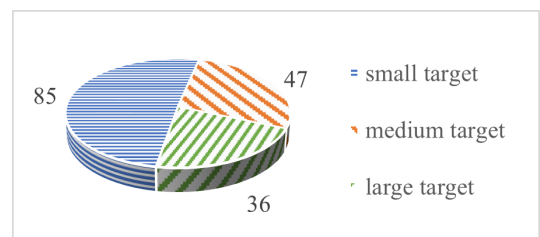


Fig. 10. Proportion of smoke objects at various scales in the test set.

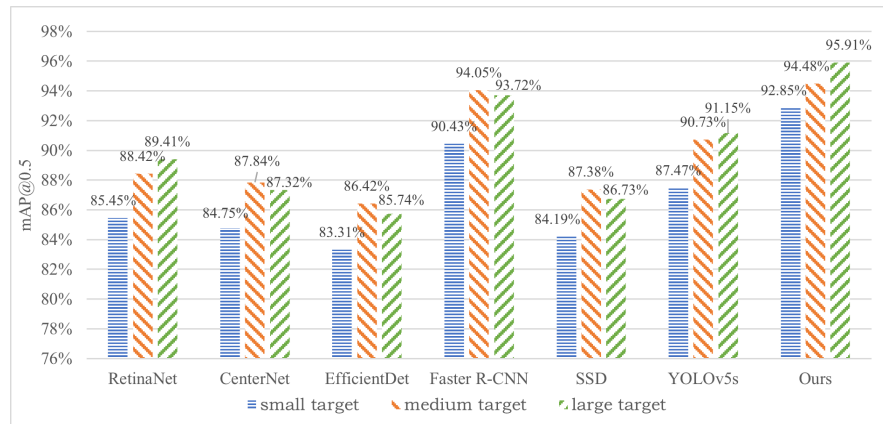


Fig. 11. Multi-scale object comparison experimental results.

Based on the multi-scale object comparison experimental results shown in Fig. 11, it can be observed that CenterNet, EfficientDet, SSD, and YOLOv5s have higher mean Average Precision (mAP) in detecting large and medium objects. However, they are slightly inferior in detecting small objects. RetinaNet and Faster R-CNN perform well in detecting objects of different scales but have relatively low overall mAP. On the other hand, our improved algorithms, especially with the optimization of light-BiFPN, achieve significantly better detection accuracy for small objects and higher mAP in object detection at various scales.

#### F. Detection Performance Analysis

The YOLOv7tiny algorithm performs poorly in detecting sparse smoke due to its low concentration in the early stages of a fire. This is because sparse smoke appears semi-transparent, often leading to false alarms [Fig. 12(a), Fig. 12(b)], missed detections [Fig. 12(d)], and low detection accuracy [Fig. 12(c)]. However, after introducing the  $\alpha$ RGB feature, our algorithm significantly improves the detection capability of sparse smoke [Fig. 13(a-d)]. Nonetheless, due to the limited proportion of low-concentration smoke in the dataset, occasional cases may arise where the detected bounding boxes do not align with the actual ones [Fig. 13(f)]. By optimizing the light-BiFPN, our algorithm achieves more accurate detection of small targets [Fig. 13(d)] and performs closer to ideal in complex environments [Fig. 13(e)].

Fig. 12 shows the detection performance of the YOLOv7tiny algorithm, while Fig. 13 depicts the detection performance of our improved algorithm.

#### V. CONCLUSION

Fire and smoke detection plays a significant role in ensuring fire safety. By combining computer vision technology to accurately locate early-stage smoke in a fire, it serves as an important tool for fire warning and prevention of fire spread. To improve the detection of small and sparse smoke in the early stages of a fire, this study extracts features related to smoke concentration, improves feature fusion structures, and optimizes algorithm complexity. By comparing with other models on a self-made dataset, the recall rate reaches 95.62%, mAP reaches 94.03%, and the detection FPS is increased

to 118.78. The algorithm complexity is reduced to 4.97M. The experimental results demonstrate the superiority of the improved algorithm in detecting sparse smoke. In future work, the algorithm will be further optimized in two aspects: firstly, by increasing the proportion of sparse smoke in the dataset to enhance algorithm robustness; secondly, by attempting to enhance algorithm expression capability through attention mechanisms, thereby further improving detection accuracy.

#### ACKNOWLEDGMENT

This work was partially supported by the Science and Technology Research Projects of Henan Province (No.222102210021), and the Key Scientific Research Project Plan for Higher Education Institutions of Henan Province, China (No.23A520004).

#### REFERENCES

- [1] "In 2021, the number of fire incidents reached a record high, with 745,000 fire extinguishments." National Fire and Rescue Administration, Jan. 2022. <https://www.119.gov.cn/gk/sj/tj/2022/26442.shtml>.
- [2] J. He, L. Li, H. Lin, and G. Xu, "Overview of Research on Smoking Detection Methods in Computer Vision." *Computer Engineering and Applications*, pp. 1-19, 2023.
- [3] 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, pp. 1-17, Feb. 2021.
- [4] A. Yazdi, H. Qin, C. Jordan, L. Yang, and F. Yan, "Nemo: An Open-Source Transformer-Supercharged Benchmark for Fine-Grained Wildfire Smoke Detection." *Remote Sensing*, vol. 14, no. 16, pp. 3979, Aug. 2022.
- [5] L. He, X. Gong, S. Zhang, L. Wang, and F. Li, "Efficient Attention Based Deep Fusion CNN For Smoke Detection in Fog Environment." *Neurocomputing*, vol. 434, pp. 224-238, Apr. 2021.
- [6] X. Sun, L. Sun, and Y. Huang, "Forest Fire Smoke Recognition Based on Convolutional Neural Network." *Journal of Forestry Research*, vol. 32, no. 5, pp. 1921-1927, Oct. 2021.
- [7] F. Wang, "Research and Implementation of Forest Fire Detection System Based on Deep Learning." *University of Electronic Science and Technology of China*, 2022.
- [8] J. Ren, W. Xiong, Z. Wu, and M. Jiang, "Fire detection and identification based on improved YOLOv3." *Computer System and Application*, vol. 28, no. 12, pp. 171-176, Dec. 2019.
- [9] C. Cao, X. Tan, X. Huang, Y. Zhang, and Z. Luo, "Study of flame detection based on improved YOLOv4." *Journal of Physics: Conference Series*, vol. 1952, no. 2, Jun. 2021.

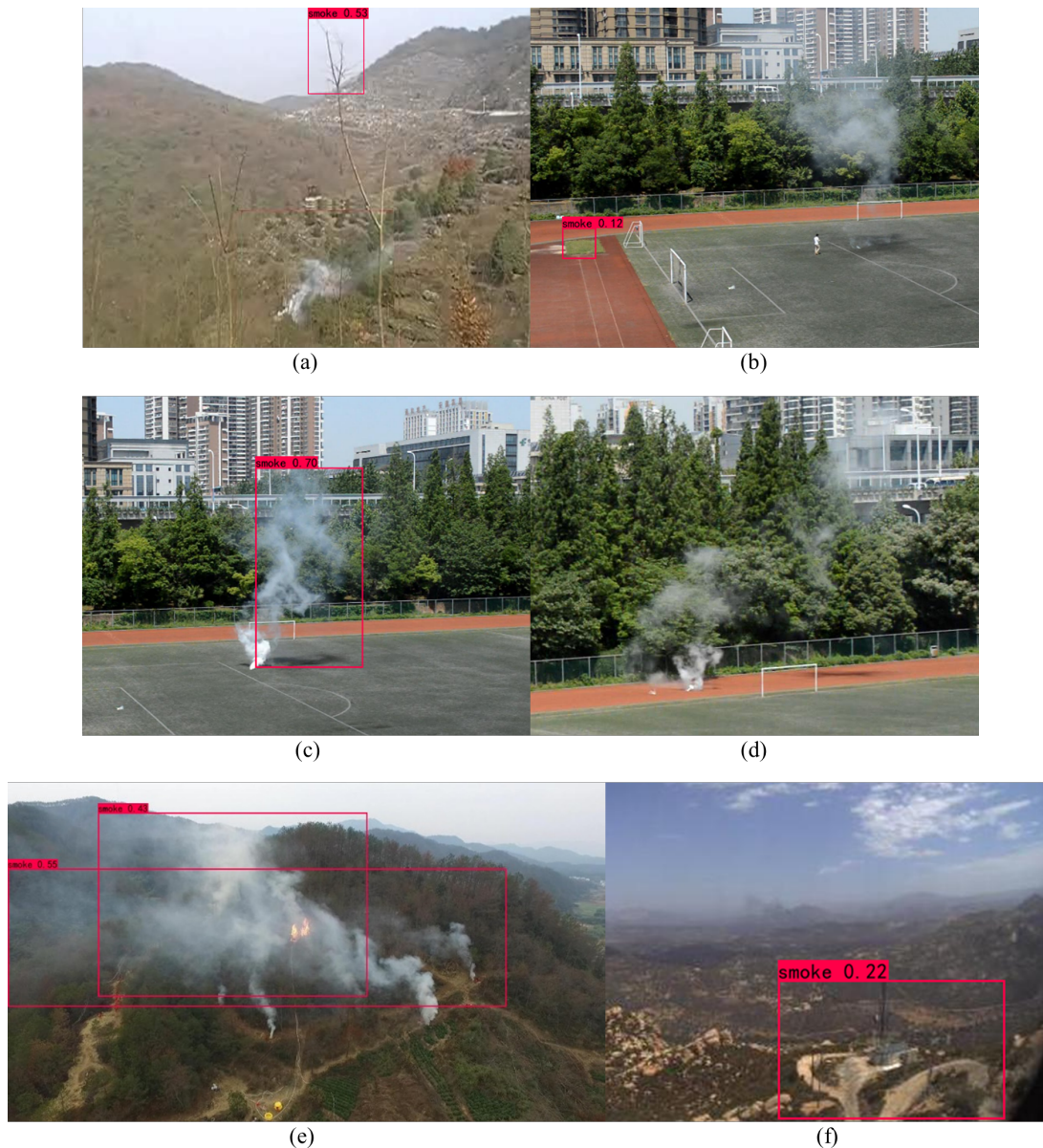


Fig. 12. YOLOv7tiny Detection Performance.

- [10] Z. Xue, H. Lin, and F. Wang, "A small target forest fire detection model based on YOLOv5 improvement." *Forests*, vol. 13, no. 8, pp. 1332, Aug. 2022.
- [11] A. Sukumaran, and T. Brindha, "Nature-inspired hybrid deep learning for race detection by face shape features." *International Journal of Intelligent Computing and Cybernetics*, vol. 13, no. 3, pp. 365-388, Aug. 2020.
- [12] T. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection." *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 936-944, 2017.
- [13] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation." *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8759-8768, 2018.
- [14] C. Wang, I. Yeh, and H. LIAO, "You Only Learn One Representation: Unified Network for Multiple Tasks." *arXiv preprint arXiv:2105.04206* (2021).
- [15] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 12, pp. 2341-2353, Dec. 2011.
- [16] H. Mo, and Z. Xie, "Smoke Concentration Measurement Method Based on Dual-Channel Deep CNN." *Pattern Recognition and Artificial Intelligence*, vol. 34, no. 9, pp. 844-852, Sep. 2021.
- [17] M. Tan, R. Pang, and Q. Le, "EfficientDet: Scalable and Efficient Object Detection." *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.10778-10787, 2020.
- [18] A. Howard, M. Sandler, B. Chen, W. Wang, L. Chen et al., "Searching for Mobilenetv3" *IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 1314-1324, 2019.
- [19] H. Zhang, M. Cisse, Y. Dauphin, and D. Lopez-Paz, "MixUp: Beyond empirical risk minimization." *6th International Conference on Learning Representations (ICLR)*, 2018.
- [20] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal Loss for Dense Object Detection." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 318-327, Feb. 2020.



Fig. 13. Improved YOLOv7tiny detection performance.

- [21] X. Zhou, J. Zhuo and P. Krähenbühl, "Bottom-Up Object Detection by Grouping Extreme and Center Points."IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 850-859, 2019.
- [22] M. Tan, R. Pang and Q. V. Le, "EfficientDet: Scalable and Efficient Object Detection."IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 10778-10787, 2020.
- [23] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks."IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), vol. 39, no. 6, pp. 1137-1149, Jun. 2017.
- [24] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A. Berg, "SSD: Single shot multibox detector."Lecture Notes in Computer Science, vol. 9905, pp. 21-37, 2016.
- [25] "yolov5."Ultralytics, 2021. <https://github.com/ultralytics/yolov5>.

# Virtual Machine Allocation in Cloud Computing Environments using Giant Trevally Optimizer

Hai-yu zhang\*

School of Information, Shanxi College of Finance and Taxation, Taiyuan, 030024, China  
School of Finance and Economics, Taiyuan University of Technology, Taiyuan 030024, China

**Abstract**—Cloud computing has gained prominence due to its potential for computational tasks, but the associated energy consumption and carbon emissions remain significant challenges. Allocating Virtual Machines (VMs) to Physical Machines (PMs) in cloud data centers, a known NP-hard problem, offers an avenue for enhancing energy efficiency. This paper presents an energy-conscious optimization approach utilizing the Giant Trevally Optimizer (GTO) which is inspired by the hunting strategies of the giant trevally, a proficient marine predator. Our study mathematically models the trevally's hunting behavior when targeting seabirds. The trevally's approach involves strategic selection of optimal hunting locations based on food availability, including pursuing seabird prey in the air or seizing it from the water's surface. Through extensive simulations, our method demonstrates superior performance in terms of skewness, CPU utilization, memory utilization, and overall resource allocation efficiency. This research offers a promising avenue for addressing the energy consumption challenges in cloud data centers while optimizing resource utilization for sustainable and cost-effective cloud operations.

**Keywords**—Cloud computing; resource allocation; virtualization; Giant Trevally Optimizer

## I. INTRODUCTION

The flexibility of cloud computing enables the provision of infrastructure, platforms, and software services. It has gained increasing popularity in private and public institutions due to its pay-per-use pricing scheme [1]. Cloud computing offers numerous advantages, such as scalability, flexibility, and cost efficiency. However, one pressing issue associated with cloud computing is its significant energy consumption [2]. Cloud data centers, which host the infrastructure and servers powering the services, consume a substantial amount of energy to handle computing tasks and store vast amounts of data. This energy-intensive operation contributes to environmental concerns, including carbon emissions and strain on power grids [3, 4]. To mitigate this problem, efforts are underway to develop energy-efficient practices such as server consolidation, virtualization, and green data center designs. By addressing the energy consumption challenge, cloud computing can become more sustainable and cost-effective while minimizing environmental impact [5]. According to the 2020, state of the data center report, data centers exhibit rack densities of 8.2 kW, with the potential to achieve 43 kW per rack through the implementation of effective water-cooling methods [6]. In the United States alone, data centers consume an estimated 140 billion kWh of energy annually [7]. In contrast, the global energy consumption of data centers is projected to range from

200 TWh to 500 TWh [8], accounting for approximately 1% of global electricity consumption. Predictions from [9] indicate that by 2030, data centers are expected to consume 3-13% of the world's electricity [10].

The convergence of Internet of Things (IoT), smart grids, meta-heuristic algorithms, machine learning, Artificial Intelligence (AI), association rule mining, and urban public transportation plays a pivotal role in revolutionizing the landscape of cloud computing. IoT sensors and devices generate an unprecedented volume of data, which smart grids harness to optimize energy distribution [11-13]. Meta-heuristic algorithms are essential for efficiently allocating resources in cloud data centers to manage this influx of data [14, 15]. Machine learning and AI algorithms analyze this data, predicting energy demands and enabling proactive resource allocation in cloud infrastructure [16-18]. Additionally, association rule mining identifies patterns and correlations within IoT-generated data, aiding in predictive maintenance and energy optimization [19]. Urban public transportation systems leverage IoT for real-time data collection and route optimization. Cloud computing serves as the backbone for processing, analyzing, and delivering information to commuters and traffic management systems, enhancing urban mobility [20].

Inefficient utilization of computing resources in cloud data centers is a significant concern that leads to excessive energy consumption. Despite the growing demand for cloud services, many data centers operate at low resource utilization levels, with an average utilization of less than 30%. This inefficiency leads to a significant amount of energy being consumed by idle nodes, accounting for more than 70% of the peak energy consumption [21]. This wastage of energy results in increased ownership costs and reduced returns on investments in cloud infrastructure. Cloud service providers increasingly recognize the importance of enhancing energy efficiency in their data centers. They are actively seeking strategies to optimize resource utilization and minimize energy wastage in order to meet the growing demand for sustainable operations. By implementing energy-efficient practices and optimizing resource allocation, they aim to achieve a more sustainable and cost-effective operation while meeting the increasing demands of cloud services [22].

This paper introduces a novel strategy for cloud computing resource allocation based on the Giant Trevally Optimizer (GTO). By adopting a cloud-based model, data can be processed, recorded, and retrieved simultaneously, ensuring efficient resource allocation. This approach optimizes resource

allocation by considering the user's request while maintaining system performance. Task assignment to virtual machines (VMs) is primarily determined by factors such as cost, deadline, and runtime. The structure of the paper is outlined as follows: Section II offers a detailed review of existing cloud resource allocation techniques. Section III elaborates on the proposed algorithm, providing details on its methodology. Section IV presents the experimental results obtained by implementing the algorithm. Finally, Section V provides a comprehensive summary of the paper and offers suggestions for future research directions in the field of cloud resource allocation.

## II. RELATED WORK

Hanini, et al. [23] introduced a novel approach that combines a virtual machine utilization scheme with a mechanism to regulate access to the virtual machine monitor for incoming requests. The number of active virtual machines is determined based on the workload, while the access control is determined by the number of requests. A mathematical model is utilized to describe the studied process and parameter values, and a power consumption model is developed and assessed. The evaluation of the proposed mechanism includes the use of numerical data to assess the quality of service (QoS) parameters. Additionally, the impact of the method on energy consumption behavior is thoroughly analyzed. The results of this analysis indicate a positive and beneficial influence of the proposed mechanism. Cloud computing brings forth various valuable services but also introduces security concerns related to user information privacy and the optimization of virtual machine allocation to enhance resource utilization. Dubey and Sharma [24] aim to address these challenges by developing a secure VM allocation algorithm based on an extended version of the Intelligent Water Drop (IWD) algorithm, which leverages natural phenomena. The implementation of their proposed algorithm was conducted using the CloudSim simulation toolkit. To evaluate its effectiveness, a comparison was performed against established VM allocation policies in the field of cloud computing. The experimental results from the simulations demonstrated that the proposed VM allocation policy outperformed existing approaches.

Samriya, et al. [25] have introduced a novel algorithm called the multi-objective Emperor Penguin Optimization (EPO) algorithm to optimize the allocation of virtual machines in a heterogeneous cloud environment, focusing on resource utilization. The proposed approach incorporates elements from the Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), and Binary Gravity Search (BGS) algorithms to ensure its suitability for virtual machines in data centers. A comprehensive evaluation of the proposed system was conducted using a JAVA simulation platform, which demonstrated its energy efficiency and significant advantages compared to other strategies. The results revealed that the EPO-based system effectively reduces energy consumption, minimizes SLA violations, and enhances QoS requirements, thereby providing a capable cloud service.

Devi and Kumar [26] have introduced a new VM allocation approach that effectively addresses SLA violation concerns and optimally allocates VMs to the most suitable hosts using the

Improved Grey Wolf Optimization (IGWO) algorithm. It considers various host characteristics, including CPU utilization and power consumption, to determine the most appropriate hosts for VM allocation. Additionally, the host's unused CPU and RAM resources are evaluated to maximize resource utilization. The experimental evaluation involved a random dataset with different virtual machines, and the proposed method was evaluated in comparison with existing methods such as ACO and Power-Aware Best Fit Decreasing (PABFD). The results demonstrate that the approach significantly minimizes the number of VM migrations, reduces SLA violations, and improves energy consumption. Consequently, the proposed VM allocation method promotes a green computing environment by consuming less power and maintaining a higher level of SLA compliance.

Xing, et al. [27] have formulated a VM allocation problem that aims to minimize the network bandwidth resources consumed by VMs and the total amount of power consumed by Physical Machines (PMs). To tackle this challenge, they propose the energy- and traffic-aware ACO (ETA-ACO) algorithm, which incorporates three innovative strategies for improved performance. The first strategy involves a two-step PM selection process that prioritizes PMs with lower power consumption and selects PMs that consume the least bandwidth. In the second strategy, VMs are arranged in descending order according to traffic demand. The third strategy generates a new solution by distributing components of optimal solutions across multiple solutions. Simulation outcomes validate the effectiveness of these three strategies in adapting ETA-ACO to the VM allocation problem. Addressing the challenging and critical issue of VM allocation for highly reliable cloud applications, Sheeba and Uma Maheswari [28] propose an improved Firefly algorithm-based approach. A K-means clustering algorithm is used to reduce migration time. Moreover, for optimal cluster selection in VM placement, adaptive PSO with the coyote optimization algorithm is applied. The suggested method is evaluated by examining the number of VMs, packet size, execution time, and transmission overhead. Under different constraints, the proposed method achieves improved performance and an optimal virtual machine placement scheme.

We have selected GTO as the basis for our research into VM allocation within cloud computing environments due to its unique and promising attributes that set it apart from other optimization algorithms. GTO draws inspiration from the hunting tactics of the giant trevally, a natural predator known for its exceptional hunting prowess in targeting seabirds and other prey. This distinctive approach to optimization allows us to model the allocation of VMs with a fresh perspective. The key benefits of GTO lie in its ability to effectively navigate complex optimization spaces, adapt to dynamic resource allocation scenarios, and converge towards superior solutions. Unlike conventional algorithms, GTO excels in its capacity to strategically select the optimal allocation locations for VMs based on factors such as food availability, mirroring the trevally's hunting strategy. Furthermore, GTO dynamically adapts to its pursuit, seizing opportunities whether they arise in the air or near the water's surface, mirroring the trevally's agile tactics. This adaptability makes it exceptionally well-suited to

the inherently dynamic and multifaceted challenges of VM allocation in cloud data centers. Moreover, GTO offers the advantage of enhanced exploration and exploitation capabilities, striking a delicate balance between exploiting known promising solutions and exploring new allocation possibilities.

### III. ENERGY-AWARE VM ALLOCATION APPROACH

A virtualization strategy based on the GTO is discussed in this section. Resource allocation in cloud environments depends on the architecture of the system, which allows different methods of access to the resources. Datacenter infrastructure can be provisioned by using a variety of methods and schemes. Fig. 1 illustrates the suggested architecture for

energy-efficient resource allocation, comprising three fundamental elements: service providers, users, and data center resource management. Users submit their requests to the cloud service provider first, and then the broker returns a response based on the user's requirements, the date line, and the operation of the resource services. The Cloud Information System (CIS) resource manager reviews the broker's request as soon as it reaches the data center, assesses its suitability, and makes the appropriate decision. Requests are accepted by CIS based on the availability of the system and are passed on to the allocation scheme to determine the global optimal solution. GTO is responsible for the initial placement of VMs as well as monitoring the solution.

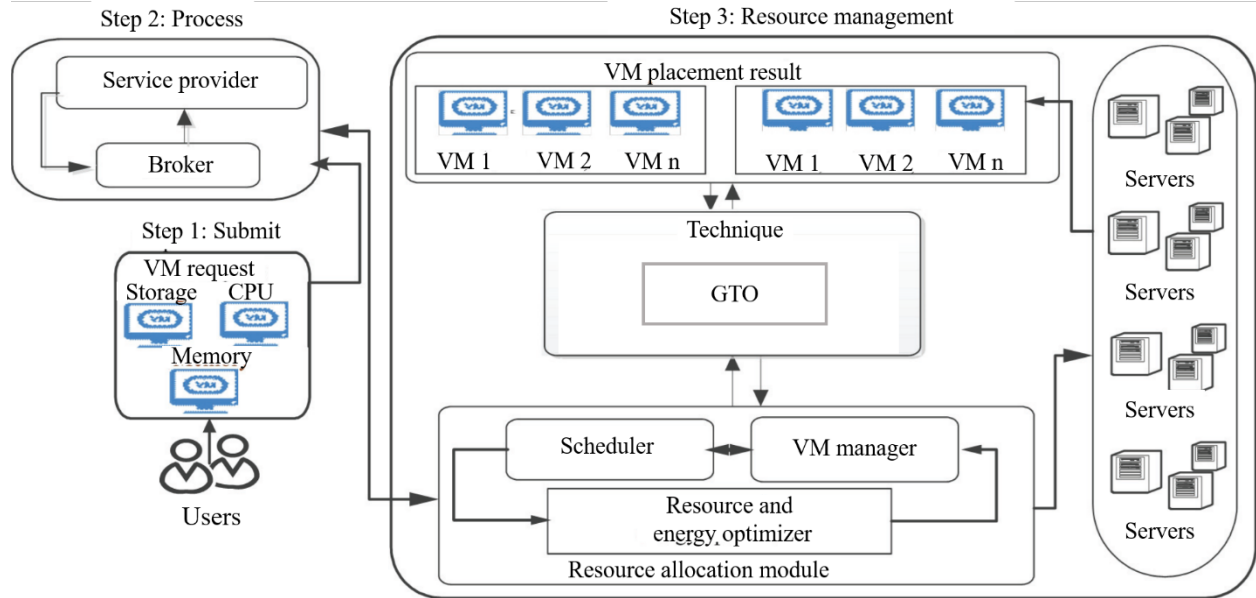


Fig. 1. Resource allocation model.

#### A. Giant Trevally Optimizer

GTO draws inspiration from nature, mimicking the behavior and strategies of giant trevallies in their pursuit of seabirds. The giant trevally belongs to the Jack family of marine predators. It is also known as the giant kingfish. The giant trevally, known as a dominant predator in its habitats, employs sophisticated hunting techniques that demonstrate its intelligence and adaptability. The giant trevally exhibits a hunting behavior that can be observed both in solitary individuals and in coordinated group efforts. It is most effective for predators to capture schooled prey when they are grouped. In a group or school, the leader, or first predator, is the most effective at capturing prey. When hunting, the giant trevally employs a remarkable strategy where it launches itself out of the water to surprise and capture its prey, often targeting seabirds.

Similarly, to other population-based meta-heuristic algorithms, GTO generates random initialization solutions termed giant trevallies. A potential or candidate solution to an optimization problem is represented by each giant trevally. These vectors, seen from a mathematical perspective as members of a population, make up the algorithm's population

matrix [29]. Eq. (1) is used to model the GTO population members.

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_i \\ \vdots \\ X_N \end{bmatrix} = \begin{bmatrix} x_{1,1} & \dots & x_{1,j} & \dots & x_{1,Dim} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{i,1} & \dots & x_{i,j} & \dots & x_{i,Dim} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ x_{N,1} & \dots & x_{N,j} & \dots & x_{N,Dim} \end{bmatrix} N \times Dim \quad (1)$$

where,  $X_i$  represents the  $i^{th}$  candidate solution of GTO,  $N$  denotes the number of GTO members,  $Dim$  denotes the number of decision parameters and  $x_{i,j}$  indicates the value of the  $j^{th}$  variable provided by the  $i^{th}$  candidate solution. When the population's size and dimensions are determined, they will not change during the experiment. Eq. (1), as originally presented, continues to serve as the foundational model for representing the GTO population members. It encapsulates the critical elements of the algorithm, wherein each giant trevally, symbolizing a potential solution, contributes to the algorithm's population matrix. Eq. (1) remains constant and integral throughout the GTO process. Every trevally in the solution space of the problem is assigned a random position prior to its operation. All feasible regions must be covered by this random assignment in the  $N \times Dim$  search space, as indicated in Eq. (2).



$$X_{i,j} = Minimum_j + (Maximum_j - Minimum_j) \times R \quad (2)$$

where,  $R$  represents a random number between 0 and 1,  $Minimum_j$  and  $Maximum_j$  indicate the limits of the described problem for the  $j^{th}$  dimension, i.e., the minimum and maximum values of population members. Each member of the GTO population is a potential solution to the VM allocation problem. Consequently, each candidate solution can be evaluated in terms of its objective function. In accordance with Eq. (3), these values are represented by a vector:

$$F = \begin{bmatrix} F_1 \\ \vdots \\ F_i \\ \vdots \\ F_N \end{bmatrix} = \begin{bmatrix} F(X_1) \\ \vdots \\ F(X_i) \\ \vdots \\ F(X_N) \end{bmatrix} N \times 1 \quad (3)$$

where,  $F_i$  refers to the  $i^{th}$  member's value of the objective function, as well as  $F$  represents the vector that contains these values.

The GTO algorithm simulates the giant trevallies' behavior while hunting for seabirds. To calculate the optimal optimization procedure of the suggested GTO algorithm, three steps are required: extensive search using Levy flight, choosing the hunting area, as well as jumping out of the water to chase and attack prey. The first and second steps represent the exploration phase of the GTO, as well as the third one represents the GTO's exploitation phase. Due to their nature, giant trevallies can travel long distances in search of food. Therefore, Eq. (4) is used in this step to simulate the foraging movements of giant trevallies.

$$X(t+1) = Best_p \times R + ((Maximum - Minimum) \times R + Minimum) \times Levy(Dim) \quad (4)$$

where  $X(t+1)$  denotes the position vector of the next-iteration giant trevally,  $Best_p$  signifies giant trevallies' current search space determined by their best position,  $R$  refers to a random number ranging from 0 to 1,  $Levy(Dim)$  stands for the Levy flight, a non-Gaussian stochastic process whose step sizes follow the Levy distribution. The algorithm is able to perform a global search due to its occasional large steps. Moreover, the levy flight increases the diversity of the population, prevents premature convergence, and enhances the ability to jump out of local optimal solutions. The recent literature has demonstrated that many animals, including marine predators, exhibit the behavior of Levy flight. Eq. (5) is used to calculate the levy (Dim).

$$Levy(Dim) = step \times \frac{u \times \sigma}{|v|^{1/\beta}} \quad (5)$$

where  $step$  refers to the step size, set to 0.01 in this case,  $\beta$  represents the Levy flight distribution index, a variable ranging from 0 to 2, set to 1.5 in this study,  $u$  as well as  $v$  correspond to random numbers normally was distributed between 0 and 1.  $\sigma$  is derived from Eq. (6).

$$\sigma = \left( \frac{\Gamma(1+\beta) \times \sin(\frac{\pi\beta}{2})}{\Gamma(\frac{1+\beta}{2}) \times \beta \times 2^{(\frac{\beta-1}{2})}} \right) \quad (6)$$

Giant trevallies determine and choose the best hunting area based on the number of food (seabirds) present in the chosen search space. This behavior is mathematically simulated by Eq. (7).

$$X(t+1) = Best_p \times A \times R + Mean_{Info} - Xi(t) \times R \quad (7)$$

where  $A$  refers to a parameter that controls position change in the range of 0.3 and 0.4,  $Xi(t)$  indicates the location of the  $i^{th}$  giant trevally in a given frame of time  $t$  (at the present iteration).  $Mean\_Info$  confirms that all of the information from the previous points has been utilized by these giant trevallies and is determined by Eq. (8).

$$Mean_{Info} = \frac{1}{N} \sum_{i=1}^N Xi(t) \quad (8)$$

Trevally starts chasing its prey during the attacking the GTO's phase. At this point, the trevally attacks the bird by jumping out of the water and catching it. During chasing and attacking prey, GTO presumed that giant trevallies experience visual distortion, which is primarily caused by the refraction of light. Refraction of light occurs as light travels from one material to another, where its direction changes at the interface. As depicted in Fig. 2, the light from point  $A$  in the first medium enters the second medium at the intersection point  $S$ . Hence refraction occurs and arrives at point  $B$  at the end of the process. The light bends toward the normal as it enters the denser medium as light travels from a rare medium, like air, to a denser medium, like water. There must be an angle between the incident and refracted rays at the point of refraction. Light rays are also affected by the medium in which they are traveling. Snell's law clarifies this connection using refractive indices, fixed values for certain media.

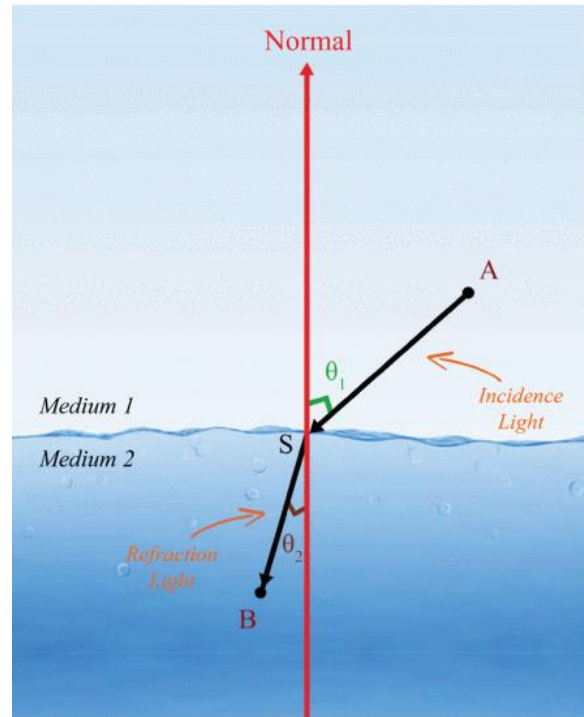


Fig. 2. Refraction of light principle.

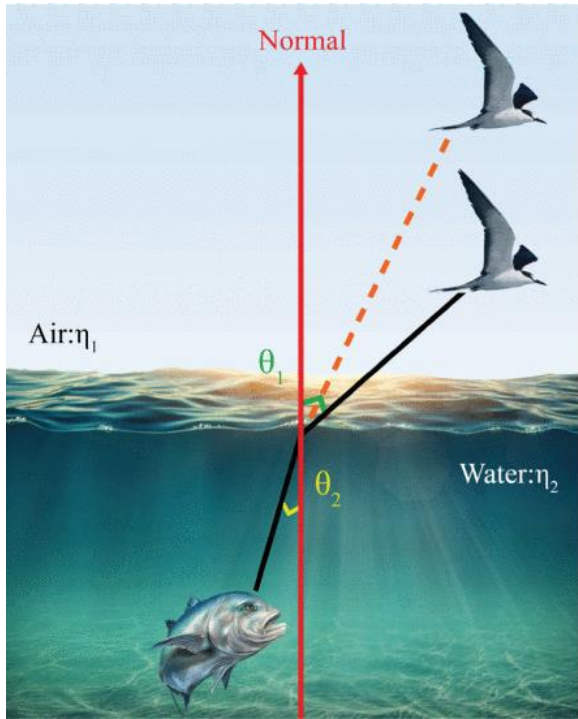


Fig. 3. Visual distortion in GTO.

As shown in Fig. 3, the giant trevally acts as an observer, as well as the bird behaves as an object. Due to the refraction of light, birds appear taller than their actual height, as indicated by the dashed line.

The relationship between the angle of incidence and the angle of refraction can be predicted using Snell's law. If we know the angle of incidence, we can determine the angle of refraction and vice versa. This relationship is demonstrated by Eq. (9), which represents Snell's law.

$$\eta_1 \sin \theta_1 = \eta_2 \sin \theta_2 \quad (9)$$

where,  $\eta_1=1.0002$  and  $\eta_2=1.3$  represent air's and water's absolute refractive indices, respectively.  $\theta_1$  and  $\theta_2$  refer to angles of incidence and refraction, respectively.  $\theta_2$  denotes a random number between 0 and 360, derived from Eq. (10).

$$\sin \theta_1 = \frac{\eta_2}{\eta_1} \sin \theta_2 \quad (10)$$

Eq. (11) is used to calculate the visual distortion.

$$v = \sin(\theta_1^\circ) \times D \quad (11)$$

where,  $\sin$  stands for the sine of a variable in degrees, and  $D$  refers to prey-attacker distance, determined by Eq. (12).

$$D = |(Best_p - Xi(t))| \quad (12)$$

where,  $Best_p$  indicates the best solution gained so far, representing the prey's location. Eq. (13) is then used to simulate giant trevally behavior during jumping as well as chasing.

$$X(t+1) = \mathcal{L} + v + \mathcal{H} \quad (13)$$

where,  $\mathcal{L}$  is the launch speed for simulating the pursuit of the bird, as determined by Eq. (14), and  $\mathcal{H}$  is the jump slope

function used by the algorithm for the adaptive transition from exploration to exploitation, derived from Eq. (15).

$$\mathcal{L} = Xi(t) \times \sin(\theta_1^\circ) \times F\_obj(Xi(t)) \quad (14)$$

$$\mathcal{H} = \mathcal{R} \times (2 - t \times \frac{2}{T}) \quad (15)$$

In Eq. (15),  $R$  stands for a random number used to denote the various motion senses of the giant trevally during the exploitation step,  $t$  signifies the current iteration, and  $T$  refers to the maximum number of iterations.

### B. User Request Model

Users request resources, commonly referred to as VMs, from the data center via a broker or cloud provider. Each resource (VM) consists of a variety of components coordinated to fulfill a certain function. UR stands for users' requests. It is possible for users to submit multiple UR requests at the same time, which are executed on a First-Come-First-Served (FCFS) basis. VMs encompass three categories of resources: storage, memory, and CPU.  $i$  and  $s$  indicate the number of resources and their measuring capacities. Eq. (16) can be used to express the request mathematically.

$$A_i \subset UR \text{ and } a_s^1, \beta_s^1, \gamma_s^1 \subset A_i$$

$$a_s^1, \beta_s^1, \gamma_s^1 \subset A_i \subset UR \Rightarrow a_s^1, \beta_s^1, \gamma_s^1 \subset UR \quad (16)$$

Eq. (17) and Eq. (18) will be used to express the request for a single resource in this case.

$$UR^1 = A_i \quad (17)$$

$$A_i = (a_s^1, \beta_s^1, \gamma_s^1) \quad (18)$$

where,  $i$  represents the number of resources required, when a user submits multiple requests, they are represented by Eq. (19) and Eq. (20).

$$UR^n = \sum_{i=1}^n A_i = A_1 + A_2 + A_3 + \dots + A_n$$

$$= (a_s^1, \beta_s^1, \gamma_s^1) + (a_s^2, \beta_s^2, \gamma_s^2) + \dots + (a_s^n, \beta_s^n, \gamma_s^n) \quad (19)$$

$$UR^n = \sum_{i=1}^n (a_s^i) + \sum_{i=1}^n (\beta_s^i) + \sum_{i=1}^n (\gamma_s^i) \quad (20)$$

### C. Resource Utilization and Energy Model

CPU and memory utilization are calculated using Eq. (21) and Eq. (22), where  $i$  represents the number of tasks assigned to  $n$  VMs.  $rpu_{ijk}$  and  $rmu_{ijk}$  represent the CPU and memory utilization of  $k$  tasks running on  $j$  VMs on the  $i^{th}$  node, respectively.

$$RPU_i = \sum_{j=1}^n \sum_{k=1}^l rpu_{ijk} \quad (21)$$

$$RMU_i = \sum_{j=1}^n \sum_{k=1}^l rmu_{ijk} \quad (22)$$

The power consumption of the  $i^{th}$  PM in terms of memory and CPU utilization can be calculated using Eq. (23), where  $t$  is the unit of time and  $C$  is the number of memory units.

$$PC_i = \frac{(RPU_i)(RMU_i)}{C} \times t \quad (23)$$

Another objective of this research is to optimize the time required to assign VMS to relevant hosts. Allocation operations

are influenced by the capacity of the hosts. A numerically generated data set is generated for each source between 0.1 and 10 milliseconds. The total allocation time is calculated by adding the CPU time associated with each host using Eq. (24).

$$Time = \sum_{i=1}^n T_i \quad (24)$$

#### D. step-by-step algorithmic explanation

The proposed VM allocation approach follows the following steps:

- Initialization: Initialize the GTO algorithm by generating a population of random solutions, referred to as "giant trevallies." Each giant trevally represents a potential solution to the VM allocation problem. Define parameters: N (number of giant trevallies), Dim (number of decision parameters), and set the population size and dimensions.
- Random position assignment: Assign each giant trevally a random position within the feasible search space of the problem, ensuring coverage across the  $N \times Dim$  search space.
- Objective function evaluation: Evaluate the objective function for each giant trevally, representing the quality of their respective VM allocation solutions. This result in a vector F containing these objective function values.
- Exploration phase: Simulate the exploration behavior of giant trevallies by employing Levy flights. This phase allows for extensive search and occasional large steps. Update the position of each giant trevally using Eq. (4), where Levy flight is used to determine the next iteration's position.
- Choosing the hunting area: Giant trevallies select their hunting areas based on the number of seabirds (food) in those areas. This is simulated using Eq. (7), which determines the new position based on a combination of the best search space and previous positions.
- Chasing and attacking prey (exploitation phase): During this phase, giant trevallies pursue and attack prey, simulating their behavior when capturing seabirds. Visual distortion is considered due to the refraction of light, which affects the perceived size of prey. This is calculated using Snell's law Eq. (9) to determine the angle of refraction. The position update during the chase is determined by Eq. (13), where L represents the launch speed, v is the visual distortion, and H is the jump slope function.
- Iteration and convergence: Repeat the above steps for a specified number of iterations or until convergence criteria are met (as defined by T, the maximum number of iterations).

#### IV. EXPERIMENTAL RESULTS

In this section, we conduct a comparison between the performance of our proposed resource allocation algorithm and previous approaches. Additionally, we perform several experiments to evaluate the effectiveness of our algorithm. The

suggested algorithm is implemented and simulated using Matlab simulator 2016b. To assess the effectiveness of the optimization algorithm, we utilize key performance indicators such as skewness, CPU utilization, memory utilization, and resource utilization. These metrics allow us to quantitatively evaluate the efficiency of our algorithm and make comparisons with other algorithms.

- Skewness: Skewness measures the asymmetry or unevenness in a probability distribution. It provides an indication of the uneven utilization of multiple resources on a server. The concept of skewness is derived from the observation that if a PM runs numerous memory-intensive virtual machines with a light load, resources may be lost due to insufficient memory to accommodate an additional virtual machine. Skewness quantifies the unevenness in resource utilization across a server by applying Eq. (25). Here,  $R$  represents the resource utilization of the  $n^{th}$  virtual machine, and  $A$  represents the average resource utilization.

$$W = \left(\frac{R_n}{A} - 1\right)^2 \quad (25)$$

- CPU utilization: This metric represents the average amount of CPU consumed by all servers while handling user requests. It is computed using Eq. (26), where  $H_i$  denotes the total number of available CPU resources and  $E_i$  represents the CPU resources requested for task execution.

$$C = \sum_{i=1}^y \frac{E_i}{H_i} \quad (26)$$

- Memory utilization: Memory utilization refers to the fraction of the memory resource that is used over time for processing all submitted tasks. It is calculated using Eq. (27), where  $v_i$  represents the total available memory and  $u_i$  indicates the memory requested for task execution.

$$M = \sum_{i=1}^y \frac{u_i}{v_i} \quad (27)$$

- Resource utilization: Resource utilization is defined as the ratio of the number of allocated resources to the total number of available resources. It provides an assessment of how effectively resources are utilized and is calculated accordingly.

$$R = \frac{c}{w} \quad (28)$$

The proposed method exhibits performance enhancements compared to existing approaches when considering 15 virtual machines. Specifically, when compared to PSO [30], genetic [31], and GWO [32] algorithms, the proposed algorithm consistently outperforms them, as depicted in Fig. 4. It achieves lower skewness values faster and maintains them even with increased iterations. This superiority is attributed to the proposed algorithm's ability to adapt swiftly and accurately to different datasets, thanks to its improved learning rate and parameter tuning. As a result, it enables more efficient optimization and better overall performance.

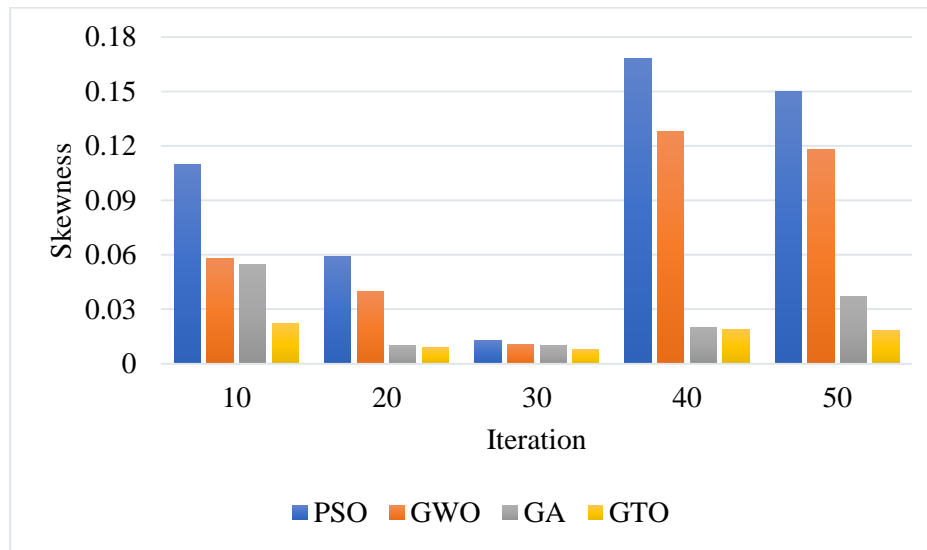


Fig. 4. Skewness comparison.

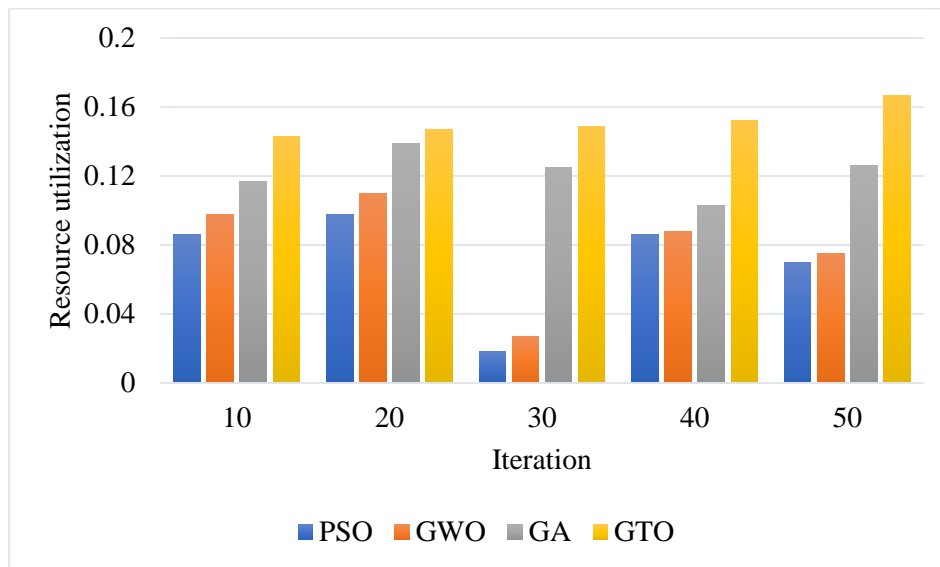


Fig. 5. Resource utilization comparison.

Fig. 5 demonstrates that the proposed algorithm utilizes more resources than existing techniques within the same number of iterations, indicating its enhanced efficiency and ability to achieve superior results with less iteration. Furthermore, Fig. 6 illustrates that the proposed algorithm exhibits improved memory utilization efficiency, requiring significantly less memory than existing techniques for the same number of iterations. Finally, Fig. 7 presents that the proposed algorithm accomplishes tasks more efficiently than existing models like GA, GWO, and PSO, as it achieves task completion in less time with the same number of iterations.

GTO in the context of VM allocation within cloud computing environments introduces a unique set of trade-offs and benefits that distinguish it from other optimization algorithms. While it may appear that GTO consumes more computational resources within the same number of iterations compared to some existing techniques, a closer examination

reveals that the unique strengths of GTO can significantly outweigh the increased resource usage, ultimately leading to improved performance, efficiency, and sustainability in various aspects of VM allocation. GTO's use of extensive search techniques, such as Levy flights, might lead to higher resource consumption in terms of computation power and time. However, this trade-off is justified by its ability to explore a wider solution space, often resulting in superior VM allocations. The increased resource usage can be considered an investment in finding more energy-efficient and effective allocation solutions. GTO strikes a balance between exploration (discovering new allocation possibilities) and exploitation (refining promising solutions). This duality is vital in tackling the NP-hard problem of VM allocation. While some algorithms might prioritize one over the other, GTO excels in both, thereby enhancing the likelihood of finding optimal or near-optimal allocations.

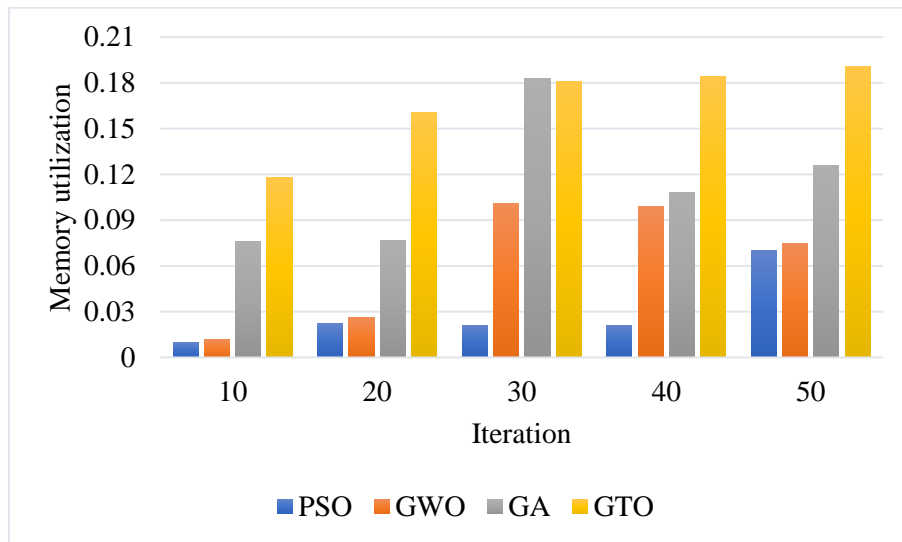


Fig. 6. Memory utilization comparison

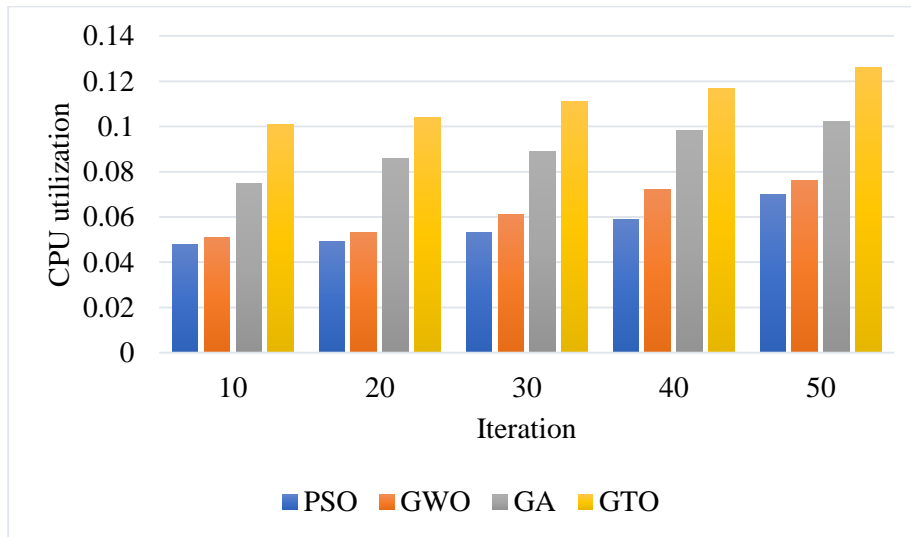


Fig. 7. CPU utilization comparison

GTO's ability to adapt its search behavior, mirroring the trevally's hunting tactics, allows it to respond effectively to changing conditions and evolving VM allocation demands. This adaptability is especially valuable in dynamic cloud environments. GTO's exploration phase, facilitated by Levy flights, enables it to perform global searches, effectively avoiding local optima. This global perspective ensures that VM allocations are not limited to suboptimal solutions, ultimately improving resource utilization and efficiency. GTO's incorporation of visual distortion due to the refraction of light is a unique feature that enhances its performance. This consideration ensures that VM allocations are not only optimal but also take into account real-world conditions, leading to more reliable and realistic allocation solutions. GTO's exploration phase increases the diversity of the population, preventing premature convergence. This diversity is crucial in avoiding stagnation and enabling the algorithm to jump out of local optima, which can be a common issue in other optimization techniques.

## V. CONCLUSION

This paper introduced an energy-conscious optimization approach based on the GTO for VM allocation. It has been compared to existing methods, including GWO, genetic, and PSO algorithms. The experimental results and performance evaluations have demonstrated the superiority of the proposed algorithm in several aspects. Firstly, the proposed algorithm consistently outperforms other algorithms in terms of skewness. It achieves lower skewness values more rapidly and maintains them even with increased iterations. This improvement is attributed to the algorithm's enhanced learning rate and parameter tuning, allowing it to adapt more effectively to different datasets. Furthermore, the proposed algorithm exhibits improved resource utilization efficiency by effectively utilizing a greater number of resources compared to existing techniques within the same number of iterations. This indicates its enhanced efficiency and ability to achieve better results with less iteration. Moreover, the algorithm demonstrates superior

memory utilization efficiency by requiring significantly less memory compared to existing techniques for the same number of iterations. This feature is valuable in resource-constrained environments where memory usage optimization is crucial. The findings highlight the promising performance and potential of the proposed GTO-based approach for VM allocation in cloud computing environments. Future research directions can explore the algorithm's applicability in different scenarios and consider additional parameters to address the complexities of diverse cloud computing environments.

While our study leverages the GTO to address VM allocation challenges in cloud computing, it is essential to acknowledge certain limitations. Firstly, GTO's resource-intensive nature, particularly in terms of computation, may pose practical constraints in real-time cloud environments where swift decision-making is crucial. Secondly, the effectiveness of the GTO algorithm may vary depending on the specific characteristics of a given cloud data center, such as size, workload, and infrastructure, which could limit its universality. Additionally, our study primarily focuses on energy efficiency and resource utilization aspects, potentially overlooking other critical performance metrics relevant to cloud service quality. Furthermore, while we account for visual distortion in VM allocation, the real-world applicability and accuracy of this consideration warrant further exploration. Despite these limitations, our research provides valuable insights into enhancing cloud sustainability and efficiency, offering a foundation for future investigations in the field.

#### 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. Iftikhar et al., "HunterPlus: AI based energy-efficient task scheduling for cloud-fog computing environments," *Internet of Things*, vol. 21, p. 100667, 2023.
- [3] Y. Kumar, S. Kaul, and Y.-C. Hu, "Machine learning for energy-resource allocation, workflow scheduling and live migration in cloud computing: State-of-the-art survey," *Sustainable Computing: Informatics and Systems*, vol. 36, p. 100780, 2022.
- [4] J. Liu, A. S. Prabuwo, A. W. Abulfaraj, S. Miniaoui, and N. Taheri, "Cognitive cloud framework for waste dumping analysis using deep learning vision computing in healthy environment," *Computers and Electrical Engineering*, vol. 110, p. 108814, 2023.
- [5] J. A. Jeba, S. Roy, M. O. Rashid, S. T. Atik, and M. Whaiduzzaman, "Towards green cloud computing an algorithmic approach for energy minimization in cloud data centers," in *Research Anthology on Architectures, Frameworks, and Integration Strategies for Distributed and Cloud Computing*: IGI Global, 2021, pp. 846-872.
- [6] P. Huang et al., "A review of data centers as prosumers in district energy systems: Renewable energy integration and waste heat reuse for district heating," *Applied energy*, vol. 258, p. 114109, 2020.
- [7] J. Ni and X. Bai, "A review of air conditioning energy performance in data centers," *Renewable and sustainable energy reviews*, vol. 67, pp. 625-640, 2017.
- [8] M. Koot and F. Wijnhoven, "Usage impact on data center electricity needs: A system dynamic forecasting model," *Applied Energy*, vol. 291, p. 116798, 2021.
- [9] A. S. Andrae and T. Edler, "On global electricity usage of communication technology: trends to 2030," *Challenges*, vol. 6, no. 1, pp. 117-157, 2015.
- [10] V. Hayyolalam, B. Pourghebleh, 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.
- [11] 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, doi: <https://doi.org/10.1016/j.suscom.2022.100822>.
- [12] 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.
- [13] T. Taami, S. Azizi, and R. Yarinezhad, "Unequal sized cells based on cross shapes for data collection in green Internet of Things (IoT) networks," *Wireless Networks*, pp. 1-18, 2023.
- [14] 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.
- [15] S. Mahmoudiazlou and C. Kwon, "A Hybrid Genetic Algorithm for the min-max Multiple Traveling Salesman Problem," *arXiv preprint arXiv:2307.07120*, 2023.
- [16] C. Han and X. Fu, "Challenge and Opportunity: Deep Learning-Based Stock Price Prediction by Using Bi-Directional LSTM Model," *Frontiers in Business, Economics and Management*, vol. 8, no. 2, pp. 51-54, 2023.
- [17] R. Soleimani and E. Lobaton, "Enhancing Inference on Physiological and Kinematic Periodic Signals via Phase-Based Interpretability and Multi-Task Learning," *Information*, vol. 13, no. 7, p. 326, 2022.
- [18] B. M. Jafari, M. Zhao, and A. Jafari, "Rumi: An Intelligent Agent Enhancing Learning Management Systems Using Machine Learning Techniques," *Journal of Software Engineering and Applications*, vol. 15, no. 9, pp. 325-343, 2022.
- [19] M. Shahin et al., "Cluster-based association rule mining for an intersection accident dataset," in *2021 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube)*, 2021: IEEE, pp. 1-6, doi: [10.1109/ICECube53880.2021.9628206](https://doi.org/10.1109/ICECube53880.2021.9628206).
- [20] S. Saeidi, S. Enjedani, E. Alvandi Behineh, K. Tehranian, and S. Jazayerifar, "Factors Affecting Public Transportation Use during Pandemic: An Integrated Approach of Technology Acceptance Model and Theory of Planned Behavior," *Tehnički glasnik*, vol. 18, pp. 1-12, 09/01 2023, doi: [10.31803/tg-20230601145322](https://doi.org/10.31803/tg-20230601145322).
- [21] 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.
- [22] V. Hayyolalam, B. Pourghebleh, M. R. Chehrehzad, and A. A. Pourhaji Kazem, "Single-objective service composition methods in cloud manufacturing systems: Recent techniques, classification, and future trends," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 5, p. e6698, 2022.
- [23] M. Hanini, S. E. Kafhali, and K. Salah, "Dynamic VM allocation and traffic control to manage QoS and energy consumption in cloud computing environment," *International Journal of Computer Applications in Technology*, vol. 60, no. 4, pp. 307-316, 2019.
- [24] K. Dubey and S. C. Sharma, "An extended intelligent water drop approach for efficient VM allocation in secure cloud computing framework," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 7, pp. 3948-3958, 2022.
- [25] J. K. Samriya, S. Chandra Patel, M. Khurana, P. K. Tiwari, and O. Cheikhrouhou, "Intelligent SLA-aware VM allocation and energy minimization approach with EPO algorithm for cloud computing environment," *Mathematical Problems in Engineering*, vol. 2021, pp. 1-13, 2021.
- [26] N. N. Devi and S. V. Kumar, "SLAV Mitigation and Energy-Efficient VM Allocation Technique Using Improved Grey Wolf Optimization Algorithm for Cloud Computing," in *2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 2022, vol. 1: IEEE, pp. 155-160.
- [27] H. Xing, J. Zhu, R. Qu, P. Dai, S. Luo, and M. A. Iqbal, "An ACO for energy-efficient and traffic-aware virtual machine placement in cloud

- computing," *Swarm and Evolutionary Computation*, vol. 68, p. 101012, 2022.
- [28] A. Sheeba and B. Uma Maheswari, "An efficient fault tolerance scheme based enhanced firefly optimization for virtual machine placement in cloud computing," *Concurrency and Computation: Practice and Experience*, vol. 35, no. 7, p. e7610, 2023.
- [29] H. T. Sadeeq and A. M. Abdulazeez, "Giant Trevally Optimizer (GTO): A Novel Metaheuristic Algorithm for Global Optimization and Challenging Engineering Problems," *IEEE Access*, vol. 10, pp. 121615-121640, 2022.
- [30] D. H. Phan, J. Suzuki, R. Carroll, S. Balasubramaniam, W. Donnelly, and D. Botvich, "Evolutionary multiobjective optimization for green clouds," in *Proceedings of the 14th annual conference companion on Genetic and evolutionary computation*, 2012, pp. 19-26.
- [31] N. Moganaragan, R. Babukarthik, S. Bhuvaneshwari, M. S. Basha, and P. Dhavachelvan, "A novel algorithm for reducing energy-consumption in cloud computing environment: Web service computing approach," *Journal of King Saud University-Computer and Information Sciences*, vol. 28, no. 1, pp. 55-67, 2016.
- [32] C. T. Joseph, K. Chandrasekaran, and R. Cyriac, "A novel family genetic approach for virtual machine allocation," *Procedia Computer Science*, vol. 46, pp. 558-565, 2015.

# Corpus Generation to Develop Amharic Morphological Segmenter

Terefe Feyisa, Dr Seble Hailu

Information Network Security Administration, Addis Ababa, Ethiopia

**Abstract**—Morphological segmenter is an important component in Amharic natural language processing systems. Despite this fact, Amharic lacks large amount of morphologically segmented corpus. Large amount of corpus is often a requirement to develop neural network-based language technologies. This paper presents an alternative method to generate large amount of morph-segmented corpus for Amharic language. First, a relatively small (138,400 words) morphologically annotated Amharic seed-corpus is manually prepared. The annotation enables to identify prefixes, stem, and suffixes of a given word. Second, a supervised approach is used to create a conditional random field-based seed-model (on the seed-corpus). Applying the seed-model (an unsupervised technique on a large unsegmented raw Amharic words) for prediction, a large corpus size (3,777,283) of segmented words are automatically generated. Third, the newly generated corpus is used to train an Amharic morphological segmenter (based on a supervised neural sequence-to-sequence (seq2seq) approach using character embeddings). Using the seq2seq method, an F-score of 98.65% was measured. Results show an agreement with previous efforts for Arabic language. The work presented here has profound implications for future studies of Ethiopian language technologies and may one day help solve the problem of the digital-divide between resource-rich and under-resourced languages.

**Keywords**—Amharic; Amharic morphology; segmentation corpus; seq2seq; under-resourced languages

## I. INTRODUCTION

Language plays a significant role in achieving the sustainable development goals (SDG 2030) [1]. Regarding language technologies, the Prime Minister of Ethiopia, Dr Abiy Ahmed, recently said, “(...) teaching Somali, Tigrinya, Amharic and Oromo languages with artificial intelligence and making these languages researchable is a great achievement (...)” [2]. This Prime Minister’s quote can be interpreted as artificial intelligence (AI) in general, and natural language processing (NLP) in particular, are issues of contemporary importance in Ethiopia. This work focuses on one aspect of Amharic NLP technology.

NLP aims at enabling machines to understand human languages. Machines usually obtain “natural language” in the form of voice or text messages. Typically, textual and vocal data are not structured and therefore require advanced technologies, like deep neural networks (DNNs), to be used and understood correctly. DNNs are mainly based on large amount of corpus for automatic feature extraction [3]. Because of this (large corpus requirement), DNNs are not being fully applied by almost all under-resourced Ethiopian languages.

Despite being the working language of Ethiopia, Amharic is one of the under-resourced languages [4], [5]. Being under-

resourced, Amharic lacks digital resources, such as sizable segmentation corpus and a morphological analyzer [6].

The lack of digital resources is mainly attributed to an expensive corpus preparation by language experts. Depending on their level of expertise, a linguist may ask starting from Ethiopian Birr 10.00 per a single word segmentation. For example, in 2019 there was a joint project (between the former Information Network Security Agency and Addis Ababa University). The aim of the project was to develop core NLP tools. The biggest share of the project cost was the payment for the linguists. Even with the minimum price, Birr 10.00 per a single word segmentation, the cost gets in millions just for about 100,000 distinct word segmentation.

The problem of language corpus scarcity is a “real-life” challenge that exists when developing NLP tools and undertaking researches.

One of the primary consequences of corpus scarcity is reflected in the approaches to develop Amharic NLP technologies. The dominant approaches to develop Amharic NLP systems are mostly rule-based, such as memory-based learning [7].

Rule-based systems have their own pros and cons [8]. Rule-based systems are advantageous, as they are declarative and are easy to comprehend, to maintain, to incorporate domain knowledge, and to trace and fix the cause of errors. However, they are heuristic and require tedious manual labor as compared to machine-learning (ML) approaches.

ML-based systems too have their own pros and cons [8]. On their advantage end, they are: trainable, adaptable, and reduces manual effort. Their disadvantages includes the requirement of: labeled corpus, retraining for domain adaptation, ML expertise to use or maintain, and they are opaque.

Given the situations, it is essential to design an alternative mechanism to enrich (with corpus), and develop language processing tools for Amharic (to make it researchable). One mechanism could be the design of an algorithm that is computationally robust and less expensive.

To design such an algorithm, a possible approach would be the use of a hybrid system: a combination of rule-based and ML-based systems. The rule-based system can be applied to generate seed-corpus for a semi-supervised ML approach as suggested by [9]: “Minimally supervised approaches provide better performance compared to applying only unsupervised methods on large unlabeled datasets.”

The purpose of this work is twofold. First, to automatically generate morph-segmented Amharic corpus. Then, the newly



generated corpus is used to develop a neural network-based Amharic morphological segmenter (AMS).

To the best of our knowledge, there has not been any work that attempted neural network techniques to develop AMS by using semi-supervised learning approach to automatically generate large amount of corpus.

The main contributions are the following:

- 1) An alternative algorithm is used to construct a morphologically segmented corpus for Amharic. The corpus is annotated with boundaries that clearly mark prefix, stem, and suffix morphemes.
- 2) A sequence-to-sequence neural network approach is used to create an Amharic morphological segmenter.
- 3) The research shall motivate the understanding of (the processing challenges of) Amharic.
- 4) The research shall inspire further research (by releasing the resources – the corpus and the algorithm – of this research to the public).

The organization of this paper is as follows: Section II provides an overview of recent advancements in morphological segmentation. Section III outlines our methodology. Subsequently, Section IV presents the results obtained from testing the method on diverse subjects. Finally, the conclusion summarizes the findings of this study and offers insights into future perspectives.

## II. RELATED WORK

### A. Why Develop Amharic Corpora

Amharic corpora is useful to develop, and apply a research work for different Amharic natural language technologies. In 2005, [6] manually annotated Amharic words (in news documents) with the most appropriate parts-of-speech (POS) tags. They managed to annotate 1,065 text documents having 202,671 words [10]. Their corpora is useful to develop probabilistic POS tagger [11] and chunker [10].

In 2016, a semi-automatic approach (very similar to this work) is followed by [12] to develop a morpho-syntactically annotated Amharic Treebank to develop a text parser. They first annotated 1,000 sentences for POS tags, morphological information, and syntactic relations of words. Using these sentences as seed-corpus, they trained a machine learning system to automatically annotate 5,000 sentences.

In 2021, [13] developed a POS tagged corpus consisting of 25,199 documents using syntactic information of words. Their corpus was tagged automatically using HornMorpho analyzer [14] with manual intervention to correct erroneous results. The morphological analyzer generates the derived stems of non-verbal words rather than basic stems. For verbs, it generates only roots rather than stems producing incorrect representations. Their corpus is not directly suitable for morphological segmentation experiments. Nevertheless, one can use their corpus as part of a seed-corpus by appropriating to a desired experiments. Regardless, HornMorpho is used by most works related to Amharic morphological segmentation [15], [16]. It is a fully-fledged morphological analysis tool for Amharic, Tigrinya and Oromo languages.

The work of [17] is also worth mentioning as, they used morphological knowledge and an extension of existing annotated dataset to improve the performance of an Amharic POS tagging system.

This paper's approach is different from HornMorpho. In that it is limited to only word segmentation task (as opposed to HornMorpho, a fully-fledged morphological analysis tool).

Brief, although morphological properties have been used to create POS tagged corpora, there is no morphologically annotated large Amharic corpus to date (i.e. that can be directly consumed by a neural network model). This study aims to construct a hybrid system to generate a morphologically annotated segmented words that can be useful for sequence-to-sequence neural network models.

### B. Segmentation for Semitic Languages

Word segmentation is regarded as a first step for almost all Semitic languages [18], [19]. This work adopts most of the methods presented for Arabic word segmenter [20]. Their method involves three steps. First, they used a small manually segmented Arabic corpus (110,000 words) to create a "seed-model". Then, they used the "seed-model" to bootstrap an unsupervised algorithm. Finally, they applied the unsupervised algorithm on a large unsegmented Arabic corpus (155 million words). They claimed a 97% exact match accuracy on a test corpus (28,449 words). A significant difference between this work and theirs is the choice of an unsupervised algorithm. Theirs is a "trigram language model". This study used a conditional random field (CRF) instead. The CRF model is a relatively better algorithm (e.g., it can use language-independent features of characters, in addition to n-grams) [18].

### C. Sequence-to-Sequence Approaches

Recently, supervised sequence-to-sequence approaches have gained success [21]–[23]. The seq2seq modeling of this work is mostly inspired (and adapts most of the techniques used) by a morphological segmentation task for the Russian language [24]. The Russian work defined MS as sequence transduction using character embeddings. They used the architecture and the hyperparameters by [22].

### D. Summary

This work builds on previous works on morphological segmentation, such as by [20]. It then enriches an already existing, manually segmented seed-corpus by applying a tool – mostly used for Amharic NLP works [14]. Finally, it adapts a seq2seq model by [24].

## III. METHODOLOGY

Amharic morphological segmentation can be modeled using only hand-crafted dataset [8]. However, building a sizable hand-crafted corpus is expensive in the amount of human work. AMS can also be designed automatically using statistical models such as Hidden Markov Model (HMM) and Conditional Random Fields (CRF) [25]. Today, one can also apply deep learning methods to get a state-of-the-art performance level [26].

This work attempts to combine the best of rule-based, ML, and deep learning approaches. To that end, it follows a three-step process (see Fig. 1).

First, a supervised approach is applied to create a **seed-model** using a hand-crafted dataset (training a CRF). Then, based on the seed-model a CRF-based unsupervised method is applied on a **raw unsegmented words** to **enrich** the manually created Amharic corpus. Third, a supervised neural **sequence-to-sequence** (seq2seq) learning approach, using character embeddings, is implemented for AMS on the enriched dataset. The seq2seq is mostly inspired by the work of [20] and the soft-attention encoder-decoder research method of [27].

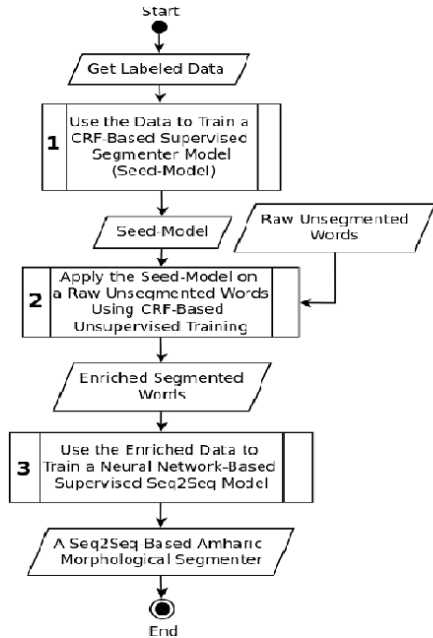


Fig. 1. A flowchart to highlight the three major sub-processes of the proposed method.

### A. Seed-Corpus for the Seed-Model

Two kinds of dataset are used to prepare a seed-model. The first one is a manually labeled, morphologically segmented corpus (173,000 words), prepared by [28].

The 173,000 segmented corpus is used for three purposes. First, it helped in generating an affixation table. Second, 80% of it (138,400 words) is used as a part of the training corpus. Third, 20% of it (34,600 words) is used as test corpus for the unsupervised CRF-based model.

The corpus by [28], however, has only representative *stems*. To compensate for the lack of *stem varieties*, another dataset, from Contemporary Amharic Corpus (CACO) by [4] is used.

Basically, the CACO corpus is a morphological analysis result of HornMorpho [14]. As such, it is not directly applica-

ble for the purpose of this study. So, it is filtered by applying a regular expression algorithm and the affixation table to get 906,417 words.

Finally, the two dataset (manually segmented corpus (138,400 words) and the filtered corpus (906,417 words)) are merged to get a total of 1,044,817 words (see Table I) as a seed-corpus to train a seed-model. All the 1,044,817 words are labeled using the “BMES” tagging scheme.

TABLE I. SUMMARY OF SEED-CORPUS PREPARATION

CORPUS	WORDS
MANUALLY SEGMENTED	138,400
FILTERED FROM CACO	906,417
MERGED TOTAL (SEED-CORPUS)	<b>1,044,817</b>

### B. The BMES Tagging Scheme

Training a word segmenter can be considered as an organized classification task with encoded classes [18].

An encoding is used to identify the presence of morph boundaries around a target character. Models using fine-grained tagging schemes contribute significantly for performance accuracy [29], [30]. As such, this work adopts the fine-grained “BMES” encoding scheme by [31]. This encoding scheme uses four class set {B, M, E, S} to capture information about the sequence of morphs in a given word. The labeling symbols have the following meanings:

- (B)egin: The start of a morph.
- (M)iddle: The continuity of a morph.
- (E)nd: The end of a morph.
- (S)ingle: Single morphs.

Table II depicts an instance of the “BMES” tagging scheme using an example of three Amharic words (/bet/ “house”, /betu/ “the house”, and /betunmko/ “and also the house”) with their corresponding manual segmentation (marked by a dash “-”) and labeling.

TABLE II. AN INSTANCE USAGE OF THE “BMES” TAG SCHEME

SEGMENTED WORD	LABELING
bet	[B, M, E]
bet-u	[B, M, E, S]
bet-u-n-m-ko	[B, M, E, S, S, S, B, E]

### C. Training a Seed-Model

The seed-corpus (1,044,817 words, labeled with the “BMES” tagging scheme) is used as an input for a supervised CRF training to prepare a seed-model. The linear-chain CRF model of the Wapiti toolkit [32] is used for both segmenting and labeling purposes. The CRF model takes a labeled seed-corpus, a template to mark n-grams and a text file to write the output (the seed-model). The seed-model’s accuracy is tested to be 96% exact match accuracy on a manually segmented test corpus of 34,600 words.

#### D. Corpus Generation from a Bulk Corpus

New corpus generation, from a bulk corpus, demands the seed-corpus, bulk-corpus and an affixation table. Having those prerequisites, a four step process follows. First, the seed-corpus is generated. Then, a bulk-corpus is prepared from the *most frequently used Amharic word lists* (347,039 **unsegmented** words) [33]. Third, a seed-model is **trained** using the seed-corpus. Finally, using the seed-model (iteratively) **re-training** is performed by using a block from the bulk-corpus. Fig. 2 is a flowchart to illustrate the process. (see also, Algorithm 1 on the following page, for detailed steps.)

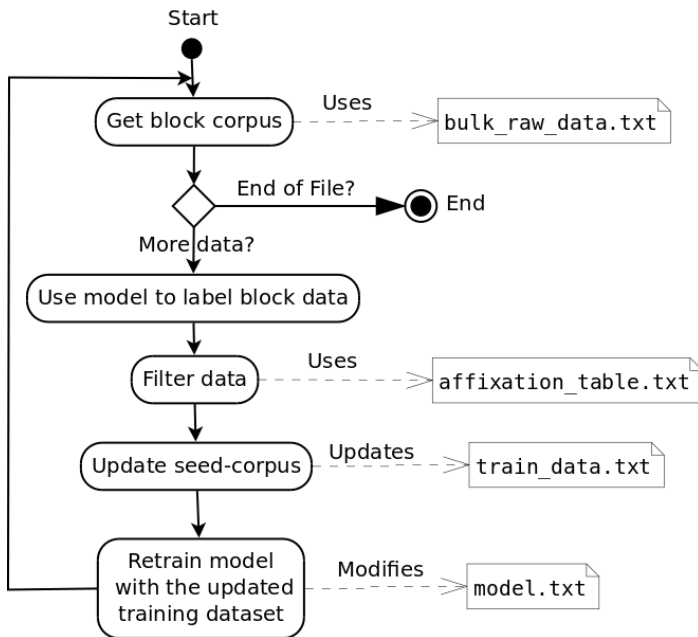


Fig. 2. Corpus preparation from a raw unsegmented corpus.

Using the algorithm, 2,732,466 segmented words are filtered from a bulk raw dataset. This corpus size (2,732,466 words) together with the seed-corpus (1,044,817) help in training the seq2seq AMS model (see Table III).

TABLE III. SUMMARY OF THE NEW CORPUS SIZE USING THE ALGORITHM

CORPUS	WORDS
SEED-CORPUS	1,044,817
CORPUS FROM BULK-CORPUS	2,732,466
MERGED DATASET FOR SEQ2SEQ	<b>3,777,283</b>

#### E. The Seq2Seq Model

The seq2seq works as transduction system [24]. That means, AMS as a seq2seq model gains an overall information from inputs and directly output a segmented sequence without using context features.

Table IV presents a sample input-output pair for a seq2seq AMS model. The input,  $\mathbf{X} = x_1, x_2, x_3, x_4$ , is the word “betu”, “the house” having 4 characters ( $x_1 = b, x_2 = e, x_3 = t, x_4 = u$ ). The output is a sequence,  $\mathbf{y}$ , having 5 characters including a boundary marker,  $\beta$ .

The final segmentation result is  $\{\text{bet}\}\beta\{\text{u}\}$ .

TABLE IV. INPUT-OUTPUT INSTANCE FOR SEQ2SEQ AMS MODEL

	Sequence	Length
<b>Input</b>	$\mathbf{X} = [b, e, t, u]$	4
<b>Output</b>	$\mathbf{y} = [b, e, t, \beta, u]$	5

The attention-based seq2seq model architecture for AMS is shown in Fig. 3. The model contains character embedding layer, an encoder layer, and an attention head and a decoder.

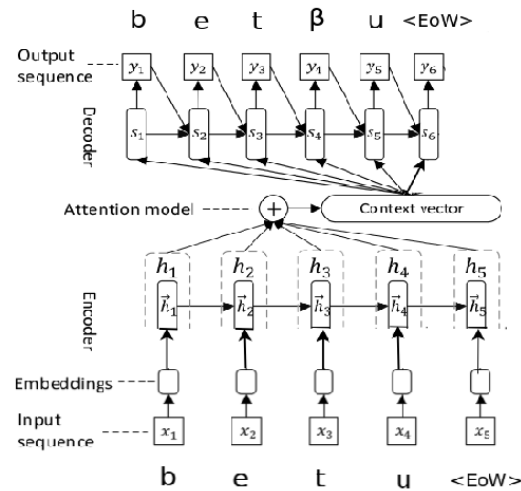


Fig. 3. Architecture of the seq2seq model (Adapted from [21]). <EoW> stands for End-of-Word.

#### F. Settings

**Dataset:** The total corpus size is 3,777,283 morphologically segmented Amharic words. This corpus is divided into two parts: 90% for training, and 10% for testing as suggested by [24]. The targets for the seq2seq model are morph-broken words (see Table V).

**Algorithm 1** Get New Corpus From Bulk Corpus

**Input:** seedCorpus, bulkCorpus, affixationTable

```

do {
    blockSize ← 1000
    begin ← 0
    end ← blockSize
    seedModel ← crfTrain(seedCorpus, 'model.txt')
    newCorpus ← seedCorpus
    while begin ≤ sizeOf(bulkCorpus)
        do {
            wordBlock ← getCorpusBlock(bulkCorpus, begin, end)
            for each word ∈ wordBlock
                do {
                    transliteratedWord ← transliterate(word)
                    writeOnFile(transliteratedWord, 'transliteratedWord.txt')
                    crfPredict('model.txt', 'transliteratedWord.txt', 'result.txt')
                    filteredSegments ← filterSegments('result.txt', affixationTable)
                    newCorpus ← merge(newCorpus, filteredSegments)
                    newCorpus ← dropDuplicates(newCorpus)
                }
            }
            newModel ← crfRetrain(newCorpus, 'model.txt')
            begin ← end+1
            end ← end + blockSize
        }
    }
return (newCorpus)

```

TABLE V. SAMPLE DATASET FOR SEQ2SEQ TRAINING AS A PAIR OF [INPUT WORD, TARGET SEGMENTED WORD]. THE EXAMPLE AMHARIC WORD HAS A ROOT **SBR**, HAVING THE SENSE OF BREAKING

INPUT WORD	SEGMENTED WORD
s babrona	s babr-o-na
s babrana	s babr-a-na
s babrwna	s babr--w-na

The Python programming language is applied for the experimentation. As for the deep learning package, Keras [34] with TensorFlow [35] as a back-end is used. For evaluating the models, the Scikit-learn [36] toolkit is used.

*G. Train the Seq2Seq Model*

The seq2seq model involves three distinct models: an encoder, an attention head and a decoder. Each of these (three) models, include a single Bidirectional Long Short Term Memory (BiLSTM) layer.

- The Long Short Term Memory (LSTM) Network  
The LSTM network architecture consists of a set of recurrently connected memory blocks, known as LSTM memory cells (dotted boxes in Fig. 4). LSTM are better at finding and exploiting long range dependencies in a data [37]. It has an input layer **X**, hidden layer **h** and output layer **y**. For example, if one inputs  $x = [b, e, t, u]$ , the house, into the LSTM, the expected prediction is a tag set as:  $y = [B, M, E, S]$  (see Fig. 4).

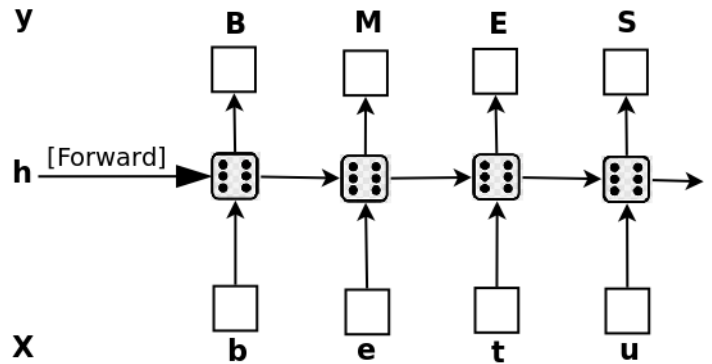


Fig. 4. The LSTM network.

- The Bidirectional LSTM  
As depicted in Fig. 5, BiLSTM is two hidden LSTM layers. In sequence tagging task, it enables us to have access to both past and future input features.

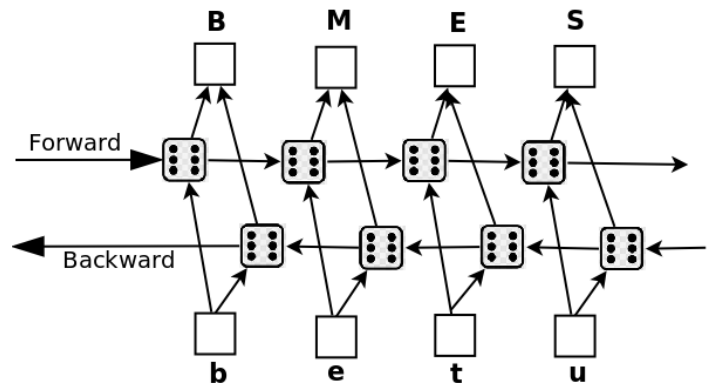


Fig. 5. A BiLSTM network.

- Hyperparameters
  - Inspired by previous works [22], [38], possible parameter combinations are explored in the preliminary experiments. The complete list of parameters is shown in Table VI. “Hidden layer size” stands for the number of BiLSTM layers or hidden state dimension and “Embeddings dimension” stands for dimensionality of the embedding layer.

TABLE VI. HYPERPARAMETERS OF THE SEQ2SEQ TRAINING.

HYPERPARAMETER	VALUE
<b>Encoder and decoder</b>	
Number of epochs	10
Number of units	1024
Batch size	64
Character embeddings	256
Optimizer	Adam
<b>Attention</b>	
Attention type	Bahdanau’s

- Training
  - The training involves three distinct models (an encoder, an attention head and a decoder, in that order) acting as a single end-to-end model.

#### Encoder

It takes a list of tokens. It converts those tokens into vectors by an embedding layer. Then, a BiLSTM layer processes the vectors sequentially. It outputs the processed sequence (for the attention head) and the internal state (useful to initialize the decoder).

#### Bahdanau’s additive attention [27]

It computes the attention weights and the context vectors.

#### Decoder

After accepting the output from the encoder, it converts the tokens into a vector using an embedding layer. The decoder keeps track of what has been generated so far using a similar layer as in the encoder. Finally, it produces context vectors and do “logit” predictions for the next token.

### H. Evaluation

Two morphological segmentation evaluation approaches are suggested by [39]. The first one is called “direct evaluation”, in which the results of a MS model are compared to “gold” standards. The other approach is known as “indirect evaluation”, where the MS models are used in other applications such as for speech recognition system.

As recommended by [24], this study uses the direct evaluation technique. So, boundary precision, boundary recall and boundary F1-score are reported.

$$\text{Precision} = \frac{\text{number of correct boundaries found}}{\text{total number of boundaries found}} \quad (1)$$

$$\text{Recall} = \frac{\text{number of correct boundaries found}}{\text{total number of correct boundaries}} \quad (2)$$

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

Where, “boundary” means the border between morphs. For example, suppose there are two boundaries in the gold standard for the Amharic word “y-bet-u” (of-the-house). If the AMS model segments this word as “y-be-t-u”, with three boundaries, one can compute precision as 67%, recall as 100% and F1-score as 80%.

## IV. EXPERIMENTAL RESULTS

### A. Results

Overall, 3,777,283 morphologically segmented Amharic words are generated with an algorithm that involves a CRF model. The CRF model’s accuracy was 96% exact match on a manually segmented test corpus of 34,600 words.

This accuracy is slightly less than that of the Arabic word segmenter by [20], which was 97% exact match accuracy on a test corpus (28,449 words). The difference may be attributed to the use of a large unsegmented Arabic corpus (155 million words) as compared to 347,039 unsegmented Amharic corpus.

Once the morphologically segmented Amharic words are generated, the next step was to develop our seq2seq model. But, before developing the seq2seq model, the newly generated data is used for training LSTM, GRU, BiLSTM, and BiGRU models in order to choose the best performing one. The performance of the models was evaluated and contrasted (see Table VII). The results showed that BiLSTM gave the best performance compared to the other models. So, we have chosen the BiLSTM model to implement our seq2seq model.

TABLE VII. A COMPARISON: TO CHOOSE THE BEST PERFORMING MODEL

MODEL	PRECISION	RECALL	F1-SCORE
GRU	91.73%	92.56%	92.14%
LSTM	92.47%	93.36%	92.91%
BiGRU	95.58%	95.95%	95.76%
<b>BiLSTM</b>	<b>98.47%</b>	<b>98.84%</b>	<b>98.65%</b>

Fig. 6 presents the attention weights for the input characters “slvbetacnmmko”, where the sound /v/, representing //, is used for technical reason. The output is the “morph broken characters” which is found to satisfy the given “gold” standard “/slv-bet-acn-n-m-ko/”.

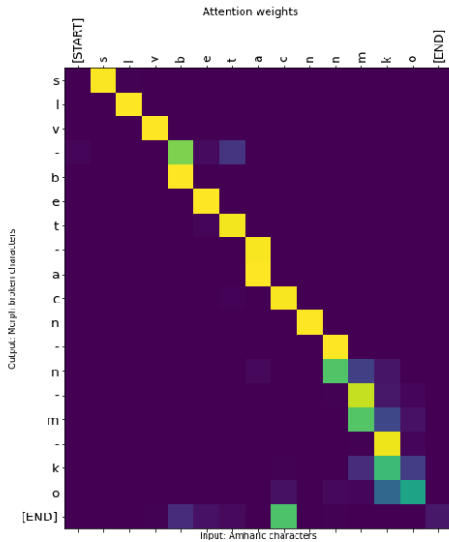


Fig. 6. Attention weights with inputs (Amharic characters) and outputs (morph broken characters) of the inputs.

### B. Discussion

For accurate ‘interpretation’ of the human language, machines should be equipped with effective natural language processing components [40]. But, this is not often the case for under-resourced languages.

Under-resourced languages are the majority of the world languages, which have not attracted much attention from researchers and donors due to economical and political reasons [40]. For these languages, corpus is a challenge to train deep learning models [41], as deep learning techniques demand large amounts of labeled corpus [3].

Amharic is one of the under resourced Ethiopian languages [4] that lacks necessary corpus and NLP applications.

To improve the situation, an Amharic morphological segmenter is implemented based on a supervised neural sequence-to-sequence approach using character embeddings, by carefully constructing language resources. But, there are still possible error sources that put our results questionable.

One possible source of error is the use of a segmentation corpus from an external morphological analyzer (HornMorpho). Errors may propagate from the morphological analyzer to the filtered corpus. However, it is difficult to spot out the exact source of errors, as this work lacks a proper error analysis.

Nevertheless, the obtained F1-Score (98.65%) indicates that, there is a window of opportunity to improve the accuracy of the Amharic morphological segmenter by applying deep neural networks.

The main focus of this work was on corpus preparation. As such, this work lacks testing and comparison of the implemented supervised neural sequence-to-sequence (seq2seq)

system against the stated previous works and with the resource-rich languages.

Comparing the obtained results with another similar implementation would have a much more impact to outline the achieved milestone. So, this can be considered as yet another limitation of this work.

One of the strongest suits of this research is the use of different datasets for seed-corpus preparation (as, having datasets for under-resourced languages is the main obstacle when it comes to the implementation of morphological segmentation systems).

### V. CONCLUSION

Unlike morphologically poor languages, such as English, Amharic language’s word segmentation resources aren’t sufficient for researchers to do their practices.

Addressing the most understudied corpus generation for a low-resource language is fascinating, and is a big step for further studies.

Using annotated datasets and unsupervised techniques, a relatively big dataset was generated. This enabled the implementation of a seq2seq-based morphological segmenter for Amharic language.

Rule-based approach is used alongside a supervised machine learning approach. This hybrid system is found to be cost-effective, flexible, and most importantly effective in constructing a language resource for segmentation.

To construct a language resource, three sources of dataset are used. The first set is 138,400 manually labeled, morphologically segmented corpus. The second source is a morphological analysis result from Contemporary Amharic Corpus (filtered using a regular expression algorithm and a rule-based method by using an affixation table) to get 906,417 words. The third source is the result of applying a corpus generator algorithm out of “*most frequently used Amharic word lists*”. Using the third technique, 2,732,466 segmented words are generated.

The newly generated segmentation corpus is then used to train a morphological segmenter model based on a supervised seq2seq neural network approach.

The seq2seq implementation involves three models appearing as one: an encoder, an attention head, and a decoder. For the seq2seq implementation, Python programming language is used on an Ubuntu machine having 64GB of memory.

The implemented seq2seq model is evaluated using a direct evaluation technique. Besides the 3,777,283 Amharic language corpus generated in the process, a 98.47% precision, a 98.84% recall, and a 98.65% F1-score have been achieved.

Brief, the major findings of this work are:

- An alternative algorithm that uses small seed-corpus to generate a large dataset from a raw bulk corpus.
- The generation of 3,777,283 morphologically segmented Amharic word corpus.
- An implementation of a seq2seq-based Amharic morphological segmenter model using the newly segmented word corpus.

- A state-of-the-art accuracy of the seq2seq morphological segmentation model (F1-score of 98.65%).

#### ACKNOWLEDGMENT

Our deep gratitude goes to Dr Derib Ado and Dr Demeke Asres Ayele who offered us valuable corpus and links. Our heartfelt appreciation goes to Dr Yemane Keleta Tedla, who sent us his full PhD dissertation paper on Tigrinya morphological segmentation.

#### REFERENCES

- [1] D. Traoré, "The role of language and culture in sustainable development," 11 2017.
- [2] N. Tube, "Dr abiy ahmed speech on artificial intelligent," 2020.
- [3] H. Liang, X. Sun, Y. Sun, and Y. Gao, "Text feature extraction based on deep learning: a review," *Eurasip Journal on Wireless Communications and Networking*, vol. 2017, 2017.
- [4] A. Mekonnen, M. Gasser, A. Nürnberger, and B. Seyoum, "Contemporary amharic corpus: Automatically morpho-syntactically tagged amharic corpus," 10 2018.
- [5] P. Rychlý and V. Suchomel, "Annotated amharic corpora," in *Text, Speech, and Dialogue* (P. Sojka, A. Horák, I. Kopeček, and K. Pala, eds.), (Cham), pp. 295–302, Springer International Publishing, 2016.
- [6] T. Yeshambel, J. Mothe, and Y. Assabie, "Morphologically annotated amharic text corpora," in *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '21*, (New York, NY, USA), p. 23492355, Association for Computing Machinery, 2021.
- [7] M. Abate and Y. Assabie, "Development of amharic morphological analyzer using memory-based learning," in *Proceedings of the 9th International Conference on Natural Language Processing (PolTAL2014)*, vol. 8686, pp. 1–13, Springer Lecture Notes in Artificial Intelligence (LNAI), 2014.
- [8] L. Chiticariu, Y. Li, and F. R. Reiss, "Rule-based information extraction is dead! long live rule-based information extraction systems!," in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, (Seattle, Washington, USA), pp. 827–832, Association for Computational Linguistics, Oct. 2013.
- [9] T. Ruokolainen, O. Kohonen, K. Sirts, S.-A. Grönroos, M. Kurimo, and S. Virpioja, "A comparative study on minimally-supervised morphological segmentation," 2015.
- [10] G. Demeke and M. Getachew, "Manual annotation of amharic news items with part-of-speech tags and its challenges," 01 2006.
- [11] M. Tachbelie, S. Abate, and L. Besacier, "Part-of-speech tagging for under-resourced and morphologically rich languages: the case of amharic," 04 2011.
- [12] B. Seyoum, E. Binyam, Y. Miyao, B. Mekonnen, and Yimam, "Morpho-syntactically annotated amharic treebank," 06 2016.
- [13] A. M. Gezmu, B. E. Seyoum, M. Gasser, and A. Nürnberger, "Contemporary amharic corpus: Automatically morpho-syntactically tagged amharic corpus," *CoRR*, vol. abs/2106.07241, 2021.
- [14] M. Gasser, "Hornmorpho 2.5 users guide," 2012.
- [15] A. T. Gebru and Y. Assabie, "Development of amharic grammar checker using morphological features of words and n-gram based probabilistic methods," in *Proceedings of the The 13th International Conference on Parsing Technologies (IWPT2013)*, pp. 106–112, 2013.
- [16] T. Dawit and Y. Assabie, "Amharic anaphora resolution using knowledge-poor approach," in *Proceedings of the 9th International Conference on Natural Language Processing (PolTAL2014)*, vol. 8686, pp. 278–289, Springer Lecture Notes in Artificial Intelligence (LNAI), 2014.
- [17] I. Gashaw and H. L. Shashirekha, "Machine learning approaches for amharic parts-of-speech tagging," *CoRR*, vol. abs/2001.03324, 2020.
- [18] Y. Tedla and K. Yamamoto, "Morphological segmentation with lstm neural networks for tigrinya," *International Journal on Natural Language Computing*, vol. 7, pp. 29–44, 04 2018.
- [19] M. Walther, "Computational nonlinear morphology with emphasis on semitic languages," *Computational Linguistics*, vol. 28, pp. 576–581, 12 2002. George Anton Kiraz (Beth Mardutho: The Syriac Institute) Cambridge: Cambridge University Press (Studies in natural language processing, edited by Branimir Boguraev and Steven Bird).
- [20] Y.-S. Lee, K. Papineni, S. Roukos, O. Emam, and H. Hassan, "Language model based arabic word segmentation," pp. 399–406, 2003.
- [21] X. Shi, H. Huang, P. Jian, Y. Guo, X. Wei, and Y.-K. Tang, "Neural chinese word segmentation as sequence to sequence translation," in *Communications in Computer and Information Science*, pp. 91–103, Springer Singapore, 2017.
- [22] D. Britz, A. Goldie, M.-T. Luong, and Q. Le, "Massive exploration of neural machine translation architectures," *arXiv preprint arXiv:1703.03906*, 2017.
- [23] T. Ruzsics and T. Samardžić, "Neural sequence-to-sequence learning of internal word structure," in *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*, (Vancouver, Canada), pp. 184–194, Association for Computational Linguistics, Aug. 2017.
- [24] N. V. AREFYEV, T. Y. GRATSIANOVA, and K. P. POPOV, "Morphological segmentation with sequence to sequence neural network," in *Komp'juternaja Lingvistika i Intelktual'nye Tehnologii*, pp. 85–95, 2018.
- [25] N. Ljubešić, "Comparing crf and lstm performance on the task of morphosyntactic tagging of non-standard varieties of south slavic languages," in *Proceedings of the Fifth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial 2018)*, pp. 156–163, 2018.
- [26] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural language processing (almost) from scratch," *Journal of Machine Learning Research*, vol. 12, no. 76, pp. 2493–2537, 2011.
- [27] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," 2014.
- [28] D. Ado, "Amharic morph order and concatenation rule." unpublished, 2021.
- [29] Y. Kitagawa and M. Komachi, "Long short-term memory for japanese word segmentation," 09 2017.
- [30] J. Yang, S. Liang, and Y. Zhang, "Design challenges and misconceptions in neural sequence labeling," *CoRR*, vol. abs/1806.04470, 2018.
- [31] T. Ruokolainen, O. Kohonen, S. Virpioja, and M. Kurimo, "Supervised morphological segmentation in a low-resource learning setting using conditional random fields," in *Proceedings of the Seventeenth Conference on Computational Natural Language Learning*, (Sofia, Bulgaria), pp. 29–37, Association for Computational Linguistics, Aug. 2013.
- [32] T. Lavergne, O. Cappé, and F. Yvon, "Practical very large scale CRFs," in *Proceedings the 48th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 504–513, Association for Computational Linguistics, July 2010.
- [33] P. Rychlý and V. Suchomel, "Annotated amharic corpora," vol. 9924, pp. 295–302, 09 2016.
- [34] Google, "Keras: The python deep learning library," 2020.
- [35] Google, "Tensorflow: An end-to-end open source machine learning platform," 2020.
- [36] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [37] J. Brownlee, "Time series prediction with lstm recurrent neural networks in python with keras," 2020.
- [38] E. Ansari, Z. Žabokrtský, M. Mahmoudi, H. Haghdoost, and J. Vidra, "Supervised morphological segmentation using rich annotated lexicon," in *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, (Varna, Bulgaria), pp. 52–61, INCOMA Ltd., Sept. 2019.
- [39] S. Virpioja, V. Turunen, S. Spiegler, O. Kohonen, and M. Kurimo, "Empirical comparison of evaluation methods for unsupervised learning of morphology," *Traitement Automatique des Langues*, vol. 52, pp. 45–90, 01 2011.

- [40] J. Muhirwe, "Towards human language technologies for under-resourced languages," 2007.
- [41] Y. Roh, G. Heo, and S. E. Whang, "A survey on data collection for machine learning: a big data-ai integration perspective," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 4, pp. 1328–1347, 2019.



# A Novel Fingerprint Liveness Detection Method using Empirical Mode Decomposition and Neural Network

Shekun Tong<sup>1\*</sup>, Chunmeng Lu<sup>2</sup>

College of Information Engineering, Jiaozuo University, Jiaozuo, Henan 454100, P. R. China<sup>1</sup>  
College of Artificial Intelligence, Jiaozuo University, Jiaozuo, Henan 454100, P. R. China<sup>2</sup>

**Abstract**—One of the most common biometric systems is fingerprint identification, which has been misused due to issues such as fraud. Hence, intelligent methods should be designed and used to recognize real-live fingerprints. Therefore, in the current work, we proposed a novel liveness fingerprint detection framework with low computational cost and excellent accuracy based on empirical mode decomposition and neural network to distinguish real from fake fingerprints. Our proposed scheme works based on empirical mode decomposition technique. The fingerprint images were cropped into  $200 \times 200$  images and then the two-dimensional (2D) images were converted into one-dimensional (1D) data, greatly reducing the computational process. The empirical mode decomposition (EMD) technique decomposed the data and the first five intrinsic mode functions (IMFs) were targeted for feature extraction through simple statistical features. The findings revealed that our suggested system can yield an average accuracy of 97.72% in distinguishing fake from real fingerprints through multilayer perceptron (MLP) neural network. This framework is very efficient compared to other techniques because only one piece of fingerprint image is enough to defend against spoof attacks. Therefore, such framework can reduce the cost of the fingerprint biometric systems, as no further hardware is needed. In addition, our framework method gives the best classification results in comparison to other previous techniques in real-live fingerprint recognition while being simple with lower computational cost. Therefore, this framework can be practically used in commercial biometric systems.

**Keywords**—Fingerprint; liveness; biometric; neural network; empirical mode decomposition

## I. INTRODUCTION

People's fingerprints have been used in criminology for many years, and today they are used in biometrics. The fingertip and its unique line pattern originate from the individual DNA pattern in each subject [1]. There are lines on the fingers of all people, which have been of interest to everyone for a long time. These important lines play different roles. One of them is to introduce frictions between finger and objects, by using this friction we can grab, write or touch objects [2]. Fingerprint is the oldest method of recognition and the progress in technology has increased its variety. One issue and difficulty in a biometric system is the lack of discrimination of fake fingerprints, to the extent that it leads to unauthorized entry into the system [3]. Hence, intelligent methods should be designed and used to recognize real-live

fingerprints. Liveness identification is an anti-spoofing technique that ensures that only the biometrics of a real and authorized individual are sent for recognition. Liveness detection relies on the fact that extra data can be collected from an authorization system, and that this extra data may be utilized to check the authenticity of an image [4]. Liveness detection utilizes either software- or hardware-based systems along with an authentication system to supply more protection. Hardware-related systems utilize more equipment and readers to capture biometric measures other than fingerprints to detect liveness. Such systems used additional equipment to record biological signals such as fingertip temperature, electric resistance, blood pressure, odor, or heartbeat [5-7]. Nixon and Rowe proposed a multispectral reader in which several light wavelengths and multiple polarizations provide extra data not available from a traditional system. According to several spectral pictures, they introduced a spoof recognition technique [8]. However, this technique has limitations due to additional hardware and remains vulnerable and unreliable. On the other side, software-related approaches utilize different image processing methods to directly process fingerprint image details for liveness detection. For example, Kiss et al. developed a hardware-related system for liveness detection, whereas Schukers et al. investigated software approaches for this purpose [9].

In general, despite the many efforts that have been made in this field, a comprehensive software-based system that is accepted by everyone has not yet been developed, and previous studies have emphasized the necessity of developing this work. Therefore, in this study, inspired by software-based systems and texture features extracted from different layers of fingerprint images, a novel feature calculation scheme was suggested using empirical mode decomposition (EMD) in a one-dimensional framework. One of the key benefits of EMD is its ability to extract hidden information from nonlinear data [10]. In the proposed method, the two-dimensional (2D) data is first converted into one-dimensional (1D) data, and then liveness is predicted through statistical features extracted from five layers of fingerprint images.

This paper is organized as follows. In Section II, various software-based solutions in the fingerprint anti-spoofing were described. Section III provides the procedure proposed in the present work. Section IV reports the experimental. Section V discusses the obtained results and Section VI makes a conclusion.

## II. RELATED WORKS

In the last two decades, many solutions have been proposed to address fingerprint spoofing vulnerabilities. Marasco et al. introduced a fingerprint liveness recognition system according to several textural properties and multiple classifiers (e.g., Bayesian classifier, decision tree, and multilayer perceptron) and achieved an accuracy of 87.5% [11]. The same authors published another paper two years later based on perspiration and morphological-based static features and reported an accuracy of 87.5% for fingerprint liveness detection [12]. Galbally et al. used image quality related features along with linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA) classifiers and reported an accuracy of 91.8% for fingerprint liveness detection [13]. Gagnaniello et al. introduced a complex liveness detection system based on Wavelet-Markov local-based features and support vector machine (SVM) and reported a good accuracy of 97.2% [14]. Nogueira et al. utilized convolutional networks with random weight and localized binary pattern along with SVM classifier and achieved an accuracy of 96.1% for liveness detection [15]. In 2015, Jiang et al. proposed co-occurrence matrix for feature extraction from fingerprint images along with SVM classifier and reported an accuracy of 93.2% for liveness detection [16]. Gottschlich et al. achieved an accuracy of 93.3% for fingerprint liveness detection using histogram of constant gradients [17]. Zhang and his colleagues used wavelet transform and localized binary patterns and reported an excellent accuracy of 97.9% [18]. Given that fingerprints show oriented texture like paradigm, Nikam et al. used Gabor filter based features to obtain local frequency and orientation data [19]. A novel feature extraction method for detecting fingerprint liveness according to the localized phase quantization has been introduced by Ghiani and his colleagues [20]. In addition, some studies have used other features such as skin deformation and fingerprint pores to detect liveness [21, 22]. For example, Espinoza and his colleagues suggested an approach through comparing pore numbers between real live and fake fingerprints [23]. Generally, previous studies show that the use of nonlinear analysis methods can achieve better classification results due to the nonlinear nature of fingerprint data. However, none of the previous researches have used the EMD method as a robust nonlinear analysis technique to extract hidden patterns in fingerprint data. Therefore, this study aims to integrate this nonlinear analysis technique with neural network in order to distinguish real fingerprints from fake ones.

## III. METHODOLOGY

In this section, the dataset used, processing algorithms and classification methods were explained in detail.

### A. Dataset

In this study, the well-known reliable database of the Liveness Detection Competition 2011 (LivDet 2011) was used that is publicly available [24]. This dataset includes 4 different subsets of fingerprint pictures captured through the Biometrika FX2000, Sagem MSO300, ItaldData ET10 and Digital Persona 4000B sensors. 4000 fingerprint images are available for every sensor. 2000 images are real-live fingerprints and the others are fake fingerprints. The fake images are synthesized by latex, gelatin, ecoflex, wood glue

and silicone. Indeed, 400 fake fingerprint images were captured for each of these five materials. Fig. 1 displays fingerprint images from the LivDet 2011 dataset.

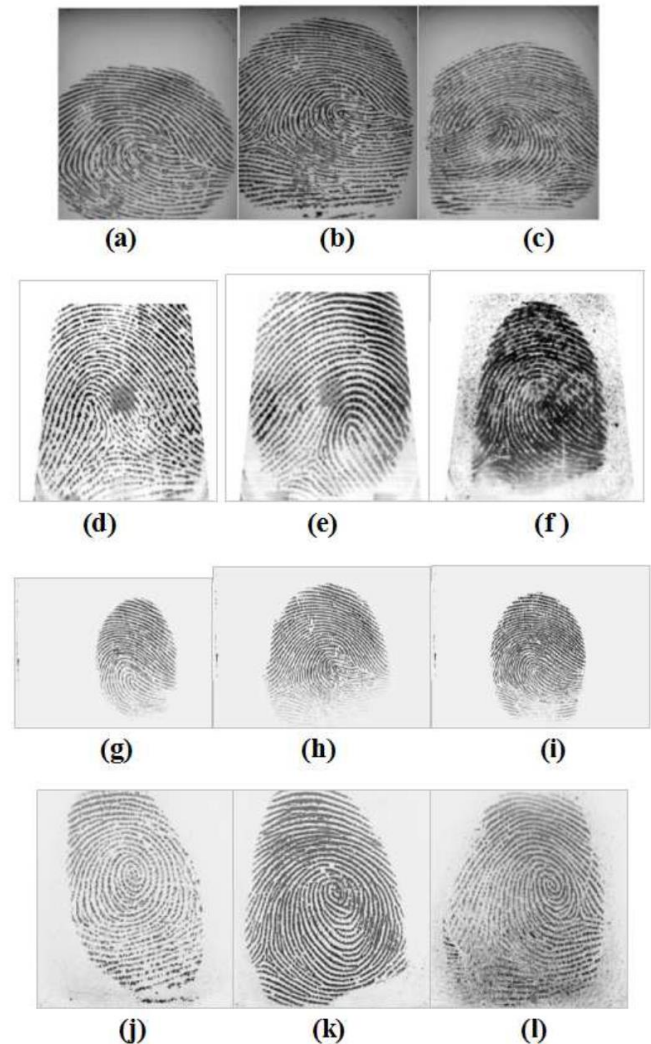


Fig. 1. Instances of spoof fingerprint pictures of the LivDet 2011 database, from Biometrika: (a) latex, (b) gelatin, (c) silicone; from Digital Persona: (d) latex, (e) gelatin, (f) silicone; from ItaldData: (g) latex, (h) gelatin, (i) silicone; from Sagem: (j) latex, (k) gelatin, (l) silicone.

### B. Preprocessing

Preprocessing is a crucial step in image processing. In the present work, images were prepared for further processing in terms of light intensity, color and other physical characteristics. This step provides a same condition for fake and real fingerprints. The actions performed in the preprocessing stage were image conversion to gray levels, image matching, image cropping, normalization, etc. Since the segments of the images were subjected to analysis, image equalization was performed after segmenting the image so that the effects of pixels around these segments do not appear in the image being processed [25]. In fact, since this work is only focused on fingerprint liveness and non-liveness, there is no need to process whole image. In addition, image cropping has two advantages: (1) analysis is performed on the fake and real fingerprint textures and the noise surrounding the picture is not processed,

and (2) processing only a small segment of the image reduces the computational cost and accelerates the processing speed. As a result, the processing system designed in this way will be more practical. Therefore, in the current research, a  $200 \times 200$  foursquare window was located on the fingerprint picture and subsequent analysis was performed on it (Fig. 2).



Fig. 2. Image segmentation and cropping used in this work.

1) *Conversion 2D data into 1D*: Since this work aimed to introduce a simple and effective system with minimum computational cost and maximum processing speed, an attempt was done to convert the 2D image data into 1D data in a simple way after cropping. This reduces the complexity of computations and simplifies the processing process. In this method, all the rows of the image pixel values matrix are sequentially placed in one row and form a vector of image pixel values. Therefore, the 2D matrix of image pixel values are transformed into a 1D vector similar to a time series, which is further processed on. This simple scheme is shown in Fig. 3.

### C. Empirical Mode Decomposition

In 1998, Huang developed a new decomposition algorithm based on the Hilbert transform called EMD. This algorithm decomposes a time series into some oscillatory signals called intrinsic mode functions (IMF) [26, 27]. Due to the ability of EMD to provide short time variations in frequency that are not attainable from Fourier transform, it may be utilized to analyze nonstationary and nonlinear signals [28]. EMD is developed in the Hilbert-Huang transform (HHT) under the supposition that each signal comprises of ordinary intrinsic functions of fluctuations [29, 30]. The nature of the algorithm is to

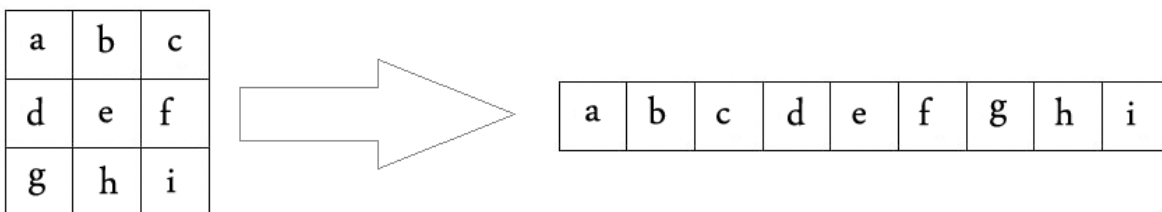


Fig. 3. A proposed scheme for converting two-dimensional data into one-dimensional data.

determine the intrinsic fluctuating functions through their characteristic temporal subscales in the signal empirically and separate it into simpler compounds correspondingly [31]. The resultant compounds obtained from the algorithm form the IMFs. IMFs are functions that satisfy two stipulates: (1) in the entire dataset, the count of extrema and the count of zero-crossings should be equal or differ not more than one; and (2) at each sample, the averaged value of the envelope determined through the local maxima and the envelope determined through the local minima approaches zero [26]. The process to produce an IMF in the EMD is known as sifting mechanism. The sifting framework to generate the IMFs of a time series  $s(t)$ , consists of the following stages:

- 1) Find all local maxima and minima of time series  $s(t)$ ;
- 2) Interpolation among the local maxima to produce lower envelope,  $s_L(t)$ , as well as interpolation among the local maxima to produce upper envelope,  $s_U(t)$ ;
- 3) For every time point  $t$ , compute the average of the lower and upper envelopes;

$$e(t) = \frac{s_L(t) + s_U(t)}{2} \quad (1)$$

- 4) Subtract the averaged resultant from the input time series;

$$d(t) = s(t) - e(t) \quad (2)$$

This is a single iteration of the sifting framework. The next stage is to verify if the time series  $d(t)$  from the previous stage is an IMF or not.

- 5) Replicate the sifting mechanism on the residue time series.

Practically, of the averaged envelop approaches zero, the sifting framework stops. This stopping condition guarantees the symmetrical property of the resultant envelop as well as the accurate relationship between the count of extremes and count of zero crossings that determine the IMFs [32].

Here, EMD was first applied to 1D data and then seven statistical features were calculated from the five first IMFs obtained from the EMD decomposition process. Previous studies on biomedical data have shown that the first five IMFs extracted from EMD contain very important details and information from the original data [33-35]. Therefore, according to previous studies and to keep the computation cost low, the first five IMFs were used in this work for feature extraction. Fig. 4 shows our proposed process according to preprocessing, EMD and feature selection approaches for fingerprint liveness identification.

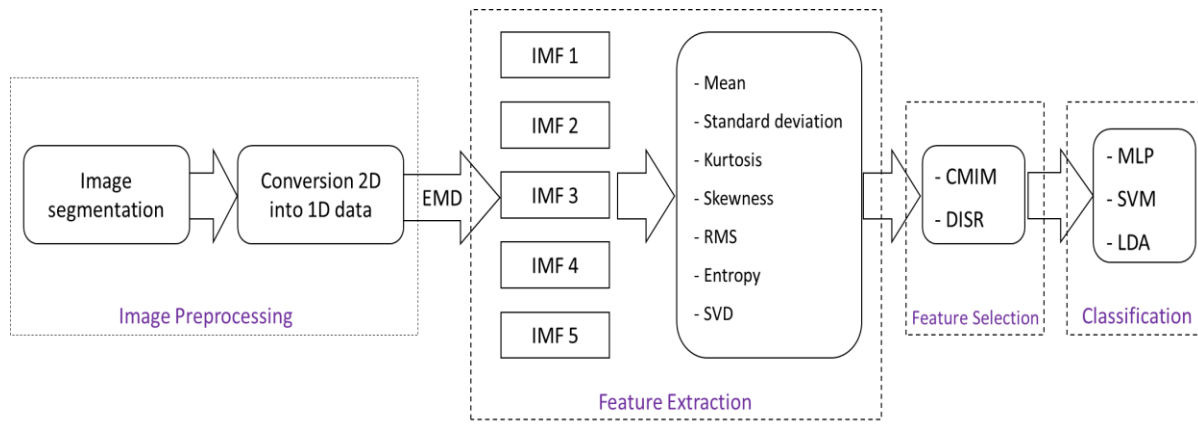


Fig. 4. The proposed process according to preprocessing, EMD and feature selection approaches for fingerprint liveness identification.

After extracting the first five IMFs for each feature vector, seven statistical features [36] (i.e., standard deviation, mean, skewness, root mean square (RMS), kurtosis, singular value decomposition (SVD), entropy) were calculated with the following mathematical definitions for each IMF:

$$\text{Mean} = \frac{1}{n} \sum_{i=1}^n (x_i) \quad (3)$$

$$\text{Standard deviation} = \left( \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{\frac{1}{2}} \quad (4)$$

$$\text{RMS} = \sqrt{\frac{1}{n} (x_1^2 + x_2^2 + x_3^2 + \dots + x_n^2)} \quad (5)$$

$$\text{Entropy} = \sum (P * \log_2 P) \quad (6)$$

$$\text{Skewness}(X) = \frac{E[(X-\mu)^3]}{\sigma^3} \quad (7)$$

$$\text{Kurtosis}(X) = \frac{E[(X-\mu)^4]}{\sigma^4} \quad (8)$$

$$\text{SVD} = U \Sigma V^T \quad (9)$$

where  $x$  or  $X$  denote the time series (i.e., each IMF),  $n$  represents the count of data points,  $P$  represents the count of image histograms,  $\mu$  denotes the mean of signal,  $\sigma$  denotes the standard deviation, and  $E[\cdot]$  is the mathematical expectation.

#### D. Feature Selection

In this work, all above seven features were calculated for the first five IMFs. Thus, a  $5 \times 7$  feature matrix was produced for every image. Therefore, 35 features were calculated for the entire fingerprint images. However, it should be noted that some features may be redundant or may not be informative for distinguishing real from fake fingerprints. Thus, the CMIM and DISR were utilized in our framework to select best discriminative features, improve the classification results and minimize computational cost.

1) *CMIM*: This method removes redundant features by making a trade-off between discrimination and independence to choose features that maximize mutual information with the class to anticipate. Conditional mutual information is defined by:

$$\text{CMI}(y; x_n | x_m) = H(y | x_n) - H(y | x_n, x_m) \quad (10)$$

Afterward, following relationship is utilized for selecting the  $(F+1)$ th feature while  $F$  features have been chosen.

$$f(F+1) = \arg \max_n (\min_{1 \leq l \leq F} I(y; x_n | x_{f(l)})) \quad (11)$$

2) *DISR*: This algorithm utilizes the following equation for feature selection [37, 38]:

$$F_{DISR} = \arg \max_{X_i \in X_s} \left\{ \sum_{X_j \in X_s} \frac{MI(X_i, X_j, Y)}{H(X_i, X_j, Y)} \right\} \quad (12)$$

where  $H(X_i, X_j, Y)$  is the information entropy and  $MI(X_i, X_j, Y)$  is the mutual information.

#### E. Classification

1) *Multilayer perceptron (MLP) neural network*: One of the simplest and effective structure of neural networks is MLP with back propagation learning procedure. MLP has been demonstrated to be effective in various problems, including pattern recognition, prediction, estimation and classification. The architecture of this neural network comprises of an input layer, hidden layer(s) and an output layer. The neurons of every layer are linked to the next layer with a certain weight, which is defined as follows:

$$\Delta W_{ij} = \eta \delta_j(n) y_i(n) \quad (13)$$

The above equation is known as the delta law through which weight correction is done from neuron  $i$  to neuron  $j$ .  $\eta$ ,  $\delta_j(n)$  and  $y_i(n)$  are learning rate variable, local gradients and input signal of neuron  $j$ , respectively. If  $j$  is a neuron in the hidden layer, then  $\delta_j(n)$  is obtained through:

$$\delta_j(n) = \varphi'_j(v_j(n)) \sum_k \delta_k(n) W_{kj}(n) \quad (14)$$

where  $k$  is a neuron in the output layer, and  $\varphi'_j(v_j(n))$  denotes the activation function for characterizing the input-output relationships of the non-linearity to entity  $j$  [39, 40].

2) *SVM*: In this study, SVM was used for classification because this classifier minimizes the expected risk in the test data and considers a margin around the class boundaries, which leads to increased generalizability of the results. SVM uses a kernel property to convert the nonlinear classification problem into a linear one by increasing the dimensionality of

the dataset. In this work, we used the RBF kernel. The mathematical notation of SVM is [41]:

$$a_i = [(\omega \cdot b_i) + x] - 1 \geq 0, \quad i = 1.2 \dots l \quad (15)$$

where  $a_i$  represents the identifier generated by SVM ( $a_i \in -1, +1$ ). This can be transformed into a dual problem via the Lagrange coefficient as follows:

$$\min Q(y) = \frac{1}{2} \sum_{i,j=1}^l y_i y_j a_i a_j \cdot K(b_i b_j) - \sum_{i=1}^l y_i \quad (16)$$

$y_i$  represents Lagrange coefficients.  $K$  is the kernel function with the following equation:

$$K(a, a_i) = \frac{\exp(-|a-a_i|^2)}{\sigma^2} \quad (17)$$

3) *LDA*: LDA is an expansion of Fisher's linear discriminant to find linear combinations of samples that separate two classes of events or objects. It is very associated with regression analysis and analysis of variance, which attempt to specify one dependent variable as a linear combination of other samples. LDA attempts to solve an optimal discrimination projection matrix  $W_{opt}$ :

$$W_{opt} = \operatorname{argmax}_W \left| \frac{W^T S_b W}{W^T S_t W} \right| \quad (18)$$

where,  $S_b$  and  $S_t$  are the scatter matrices with the following definitions:

$$S_b = \sum_{p=1}^q n_p (\mu_p - \mu)(\mu_p - \mu)^T \quad (19)$$

$$S_t = \sum_{p=1}^q n_p (\mu_p - \mu)(\mu_p - \mu)^T + \sum_p (x_p - \mu_{k_p})(x_p - \mu_{k_p})^T \quad (20)$$

where,  $S_b$  represents the between-class dispersion matrix and  $S_t$  represents the total dispersion matrix. The second term

in (20) represents the within-class dispersion matrix.  $\mu_p$  represents the averaged feature vector of image class  $p$ , as well as  $n_p$  denotes the count of features in image class  $p$ .  $q$  denotes the total count of the features.  $x_p$  denotes the feature vector of a data point, and  $\mu_{k_p}$  denotes the vector of the image class that  $x_p$  belongs to [42].

#### IV. RESULTS

After preprocessing and cropping the fingerprint images, the 2D data of the images were converted into 1D data, and then the EMD algorithm was applied to this 1D data, and the IMFs of each data were extracted for real and fake fingerprints. Next, all seven mentioned features were calculated for the first five IMFs. Fig. 5, 6, 7 and 8 show the box plots for the mean, standard deviation, entropy and SVD features for the five IMFs of real and fake fingerprints, respectively.

As shown in the above figures, there are obvious differences in the features extracted from different IMFs between real and fake fingerprints. However, as expected, the rate of change decreases after IMF1. This observation is due to the fact that the IMF 1 has more information and details from the original data, and in the subsequent IMFs, the amount of these details decreases accordingly.

After feature extraction, feature selection was performed with CMIM and DISR methods. Then, feature classification by three different classifiers (i.e., MLP, SVM and LDA) was performed to distinguish real from fake fingerprints. At this stage, 70% of the dataset (i.e., extracted features) was allocated for training classifiers, 10% of the dataset for validation, and the remaining 20% for testing the performance of the classifiers. To assess the classification performance, a random subsampling technique was used that replicates the hold-out cross validation  $n$  times. MLP training process was stopped if 1000 iterations executed or error reached less than 0.01%.

Mean of IMFs

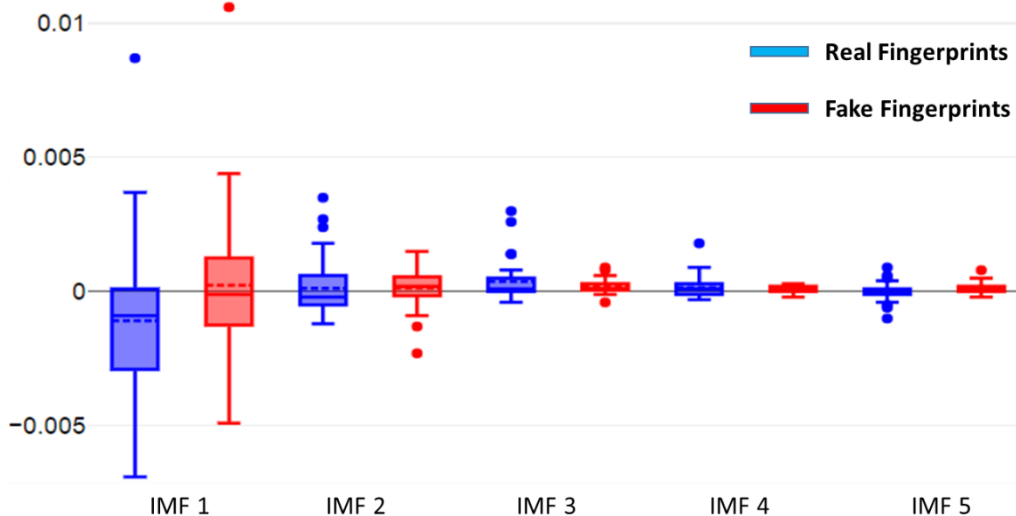


Fig. 5. Mean of the first 5 IMFs computed from one-dimensional data of real and fake fingerprints.

### Standard Deviation of IMFs

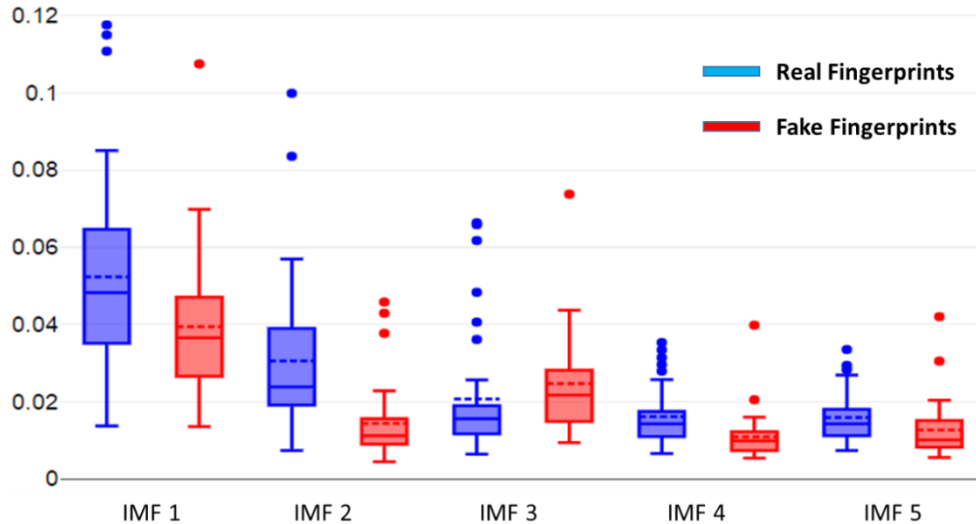


Fig. 6. Standard deviation of the first 5 IMFs computed from one-dimensional data of real and fake fingerprints.

### Entropy of IMFs

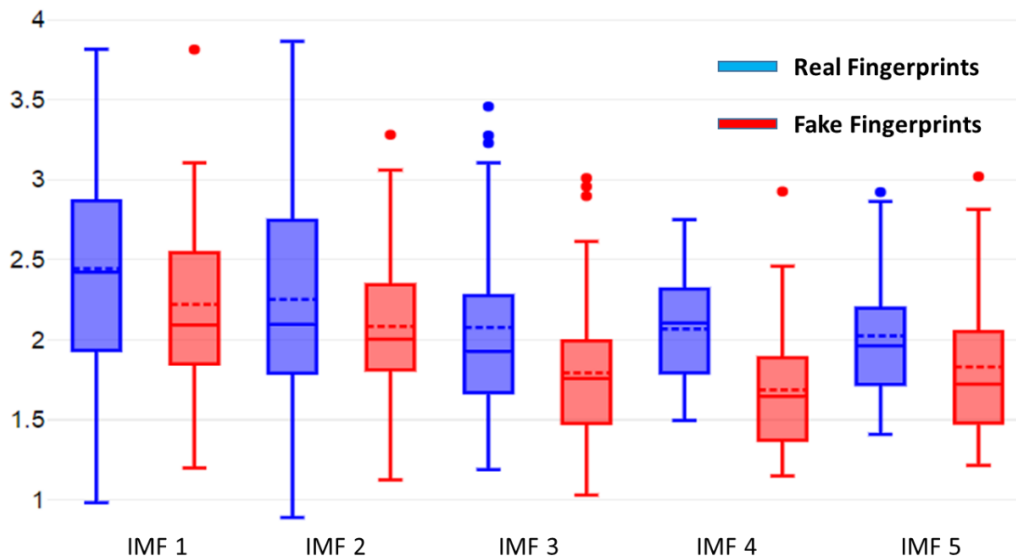


Fig. 7. Entropy of the first 5 IMFs computed from one-dimensional data of real and fake fingerprints.

To evaluate the classification performance, the following error criteria were calculated and used:

False fake rate (FFR) = The number of fake fingerprints that are mistakenly recognized as real.

False real rate (FRR) = The number of real fingerprints that are mistakenly recognized as fake.

Average classification error (ACE) = (FFR + FRR)/2.

Tables I, II and III summarize the classification results obtained by MLP, SVM and LDA classifiers, respectively. All classifiers produced a lower ACE through the features chosen by the DISR feature selection approach as input. Also, all

classifiers produced a higher ACE using all features as input. This showed that feature selection is an effective approach to feed classifiers with high discriminative features. Our experiments showed that DISR is a more effective feature selection method than CMIM, which can lead to better classification results. The best results of FFR, FRR and ACE obtained by MLP were 2.48%, 2.08% and 2.28%, respectively (Table I).

Also, the best results of FFR, FRR and ACE obtained by SVM were 3.40%, 2.24% and 2.82% respectively (Table II). Finally, the best results of FFR, FRR and ACE obtained by LDA were 5.53%, 2.64% and 4.09% respectively (Table III).

## Singular Value Decomposition of IMFs

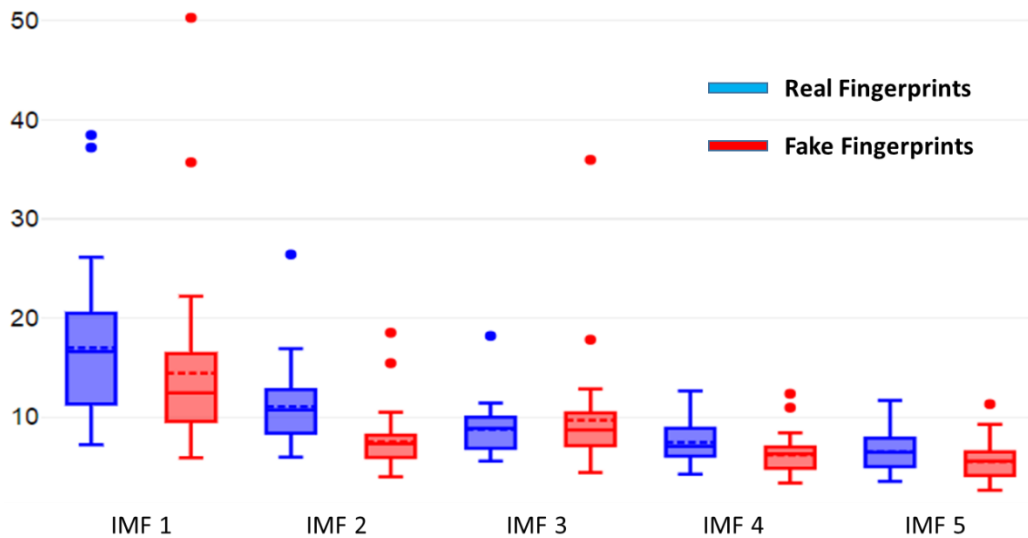


Fig. 8. SVD of the first 5 IMFs computed from one-dimensional data of real and fake fingerprints.

TABLE I. PERFORMANCE OF MLP NEURAL NETWORK IN REAL AND FAKE FINGERPRINT CLASSIFICATION

Feature set	FFR (%)	FRR (%)	ACE (%)
All features	2.55	3.30	2.93
Selected features by CMIM	2.57	2.80	2.69
Selected features by DISR	2.48	2.08	2.28

TABLE II. PERFORMANCE OF SVM CLASSIFIER WITH RBF KERNEL IN REAL AND FAKE FINGERPRINT CLASSIFICATION

Feature set	FFR (%)	FRR (%)	ACE (%)
All features	4.51	3.51	4.01
Selected features by CMIM	3.20	3.10	3.16
Selected features by DISR	3.40	2.24	2.82

TABLE III. PERFORMANCE OF LDA CLASSIFIER WITH IN REAL AND FAKE FINGERPRINT CLASSIFICATION

Feature set	FFR (%)	FRR (%)	ACE (%)
All features	6.68	5.72	6.20
Selected features by CMIM	4.95	4.45	4.70
Selected features by DISR	5.53	2.64	4.09

## V. DISCUSSION

Spoof attacks with non-real replications substantially threaten the security of different fingerprint identification systems. Thus, it is necessary to develop efficient countermeasures against these deceive attacks. In the current work, a novel liveness fingerprint detection framework with low computational cost and excellent accuracy was proposed. Our proposed scheme works based on empirical mode decomposition technique. The fingerprint images were cropped into  $200 \times 200$  images and then converted the 2D images into 1D data, greatly reducing the computational process. The EMD technique decomposed the data and the first five IMFs were targeted for feature extraction through simple statistical features. Consistent with previous studies [15, 16], our findings also demonstrated the efficacy of textural features to detect fingerprint viability. The findings revealed that our suggested

system can yield an average accuracy of 97.72% in distinguishing fake from real fingerprints through MLP neural network. This framework is very efficient compared to other techniques because only one piece of fingerprint image is enough to defend against spoof attacks. Therefore, such framework can reduce the cost of the fingerprint biometric systems, as no further hardware is needed. Image cropping, 2D to 1D data conversion and the use of nonlinear EMD analysis are the innovations of this study that distinguish our work from previous studies. As will be explained in the next paragraph, this framework led to the improvement and development of previous results and was a step forward in the development of software-based methods for fingerprint liveness detection.

In this section, our proposed framework was compared with other techniques examined on the same database (i.e., LivDet 2011 database). Table IV indicates the characteristics

and results of similar papers conducted on the LivDet 2011 database to discriminate fake from real fingerprints in terms of ACE. As indicated, seven papers have worked on this dataset with various computational algorithms to detect real-live fingerprints. Most of the previous techniques utilized texture features and all of them utilized SVM classifier. Gragnaniello et al. [43] reported the best classification results with the ACE

= 5.7%. Their system works based on local contrast phase descriptor. As shown in Table IV, our introduced technique gives the best classification results compared to other previous methods in real-live fingerprint recognition while being simple with lower computational cost. Therefore, this framework can be practically used in commercial biometric systems.

TABLE IV. CHARACTERISTICS AND RESULTS OF SIMILAR STUDIES CONDUCTED ON THE LIVDET 2011 DATASET TO DISTINGUISH REAL FROM FAKE FINGERPRINTS

Author (year)	Algorithm	Classifier	Result
Nogueira et al. (2014) [15]	Convolutional network with random weight and local binary pattern	SVM	ACE = 6.5%
Jian et al. (2015) [16]	Co-occurrence matrix	SVM	ACE = 11%
Jia et al. (2014) [44]	Multiscale local binary pattern	SVM	ACE = 7.5%
Gragnaniello et al. (2015) [43]	Local contrast phase descriptor	SVM	ACE = 5.7%
Jia et al. (2013) [45]	Multiscale local ternary patterns	SVM	ACE = 9.8%
Zhang et al. (2014) [18]	Wavelet transform and local binary patterns	SVM	ACE = 12.5%
Johnson et al. (2014) [46]	Pore characteristics	SVM	ACE = 12%
Our proposed system	Empirical mode decomposition and statistical features	MLP, SVM, LDA	ACE = 2.28%

## VI. CONCLUSION

In summary, the proposed framework includes preprocessing along with image cropping incorporation, feature extraction using nonlinear analysis, feature selection by two different information-based approach, and classification stage through neural network, improved accuracy of previous techniques for fingerprint liveness detection. The findings of the present study support the use of nonlinear analysis and texture features for liveness fingerprint detection. However, the results of this study need to be validated by additional databases. In addition, future studies should explore other advanced classification techniques, especially deep learning models, to improve our findings.

## REFERENCES

[1] Maltoni, D., et al., Handbook of fingerprint recognition. Vol. 2. 2009: Springer.

[2] Adam, D.E.E.B. and P. Sathesh, Evaluation of fingerprint liveness detection by machine learning approach-a systematic view. Journal of IoT in Social, Mobile, Analytics, and Cloud, 2021. 3(1): p. 16-30.

[3] Yuan, C., X. Sun, and R. Lv, Fingerprint liveness detection based on multi-scale LPQ and PCA. China Communications, 2016. 13(7): p. 60-65.

[4] Ghiani, L., et al., Review of the fingerprint liveness detection (LivDet) competition series: 2009 to 2015. Image and Vision Computing, 2017. 58: p. 110-128.

[5] Khaleghi, A., et al., New ways to manage pandemics: Using technologies in the era of covid-19: A narrative review. Iranian journal of psychiatry, 2020. 15(3): p. 236.

[6] Nigeria, Y.L., Analysis, design and implementation of human fingerprint patterns system "towards age & gender determination, ridge thickness to valley thickness ratio (RTVTR) & ridge count on gender detection. International Journal of Advanced Research in Artificial Intelligence, 2012. 1(2).

[7] George, J.P., S. Abhilash, and K. Raja, Transform domain fingerprint identification based on DTCWT. International Journal of Advanced Computer Science and Applications, 2012. 3(1).

[8] Ametefe, D., et al., Fingerprint liveness detection schemes: A review on presentation attack. Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization, 2022. 10(2): p. 217-240.

[9] Kavita, K., G.S. Walia, and R. Rohilla. A contemporary survey of unimodal liveness detection techniques: Challenges & opportunities. in 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS). 2020. IEEE.

[10] Mandic, D.P., et al., Empirical mode decomposition-based time-frequency analysis of multivariate signals: The power of adaptive data analysis. IEEE signal processing magazine, 2013. 30(6): p. 74-86.

[11] Marasco, E. and C. Sansone. An anti-spoofing technique using multiple textural features in fingerprint scanners. in 2010 IEEE workshop on biometric measurements and systems for security and medical applications. 2010. IEEE.

[12] Marasco, E. and C. Sansone, Combining perspiration-and morphology-based static features for fingerprint liveness detection. Pattern Recognition Letters, 2012. 33(9): p. 1148-1156.

[13] Galbally, J., et al., A high performance fingerprint liveness detection method based on quality related features. Future Generation Computer Systems, 2012. 28(1): p. 311-321.

[14] Gragnaniello, D., et al., Wavelet-Markov local descriptor for detecting fake fingerprints. Electronics Letters, 2014. 50(6): p. 439-441.

[15] Nogueira, R.F., R. de Alencar Lotufo, and R.C. Machado. Evaluating software-based fingerprint liveness detection using convolutional networks and local binary patterns. in 2014 IEEE workshop on biometric measurements and systems for security and medical applications (BIOMS) Proceedings. 2014. IEEE.

[16] Jiang, Y. and X. Liu, Spoof fingerprint detection based on co-occurrence matrix. International Journal of Signal Processing, Image Processing and Pattern Recognition, 2015. 8(8): p. 373-384.

[17] Gottschlich, C., et al. Fingerprint liveness detection based on histograms of invariant gradients. in IEEE international joint conference on biometrics. 2014. IEEE.

[18] Zhang, Y., et al. Fake fingerprint detection based on wavelet analysis and local binary pattern. in Biometric Recognition: 9th Chinese Conference, CCBR 2014, Shenyang, China, November 7-9, 2014. Proceedings 9. 2014. Springer.

[19] Nikam, S.B. and S. Agarwal, Curvelet-based fingerprint anti-spoofing. Signal, Image and Video Processing, 2010. 4(1): p. 75-87.

[20] Ghiani, L., G.L. Marcialis, and F. Roli. Fingerprint liveness detection by local phase quantization. in Proceedings of the 21st international conference on pattern recognition (ICPR2012). 2012. IEEE.

[21] Antonelli, A., et al., Fake finger detection by skin distortion analysis. IEEE Transactions on Information Forensics and Security, 2006. 1(3): p. 360-373.



- [22] Kulkarni, S.S. and H.Y. Patil, Survey on fingerprint spoofing, detection techniques and databases. *International Journal of Computer Applications*, 2015. 975: p. 8887.
- [23] Espinoza, M. and C. Champod. Using the number of pores on fingerprint images to detect spoofing attacks. in *2011 International Conference on Hand-Based Biometrics*. 2011. IEEE.
- [24] Yambay, D., et al. LivDet 2011—Fingerprint liveness detection competition 2011. in *2012 5th IAPR international conference on biometrics (ICB)*. 2012. IEEE.
- [25] Aslan, M.F., K. SABANCI, and A. Durdu, Comparison of Contourlet and Time-Invariant Contourlet Transform Performance for Different Types of Noises. *Balkan Journal of Electrical and Computer Engineering*, 2019. 7(4): p. 399-404.
- [26] Barbosh, M., P. Singh, and A. Sadhu, Empirical mode decomposition and its variants: A review with applications in structural health monitoring. *Smart Materials and Structures*, 2020. 29(9): p. 093001.
- [27] Pan, J. and Y. Tang, Texture classification based on bidimensional empirical mode decomposition and local binary pattern. *International Journal of Advanced Computer Science and Applications*, 2013. 4(9).
- [28] de Souza, U.B., J.P.L. Escola, and L. da Cunha Brito, A survey on Hilbert-Huang transform: Evolution, challenges and solutions. *Digital Signal Processing*, 2022. 120: p. 103292.
- [29] Khaleghi, A., et al., A neuronal population model based on cellular automata to simulate the electrical waves of the brain. *Waves in Random and Complex Media*, 2021: p. 1-20.
- [30] Khaleghi, A., et al., Possible Neuropathological Mechanisms Underlying the Increased Complexity of Brain Electrical Activity in Schizophrenia: A Computational Study. *Iranian Journal of Psychiatry*, 2023: p. 1-7.
- [31] Boronov, V. and V. Ompokov. The Hilbert-Huang Transform for biomedical signals processing. in *2014 International conference on computer technologies in physical and engineering applications (ICCTPEA)*. 2014. IEEE.
- [32] Zeiler, A., et al. Empirical mode decomposition-an introduction. in *The 2010 international joint conference on neural networks (IJCNN)*. 2010. IEEE.
- [33] Karagiannis, A. and P. Constantinou, Noise-assisted data processing with empirical mode decomposition in biomedical signals. *IEEE Transactions on information technology in biomedicine*, 2010. 15(1): p. 11-18.
- [34] Schiecke, K., et al., Assignment of empirical mode decomposition components and its application to biomedical signals. *Methods of information in medicine*, 2015. 54(05): p. 461-473.
- [35] Yousefi Rizi, F., A review of notable studies on using Empirical Mode Decomposition for biomedical signal and image processing. *Signal Processing and Renewable Energy*, 2019. 3(4): p. 89-113.
- [36] Khaleghi, A., et al., Applicable features of electroencephalogram for ADHD diagnosis. *Research on Biomedical Engineering*, 2020. 36: p. 1-11.
- [37] Khaleghi, A., et al., EEG classification of adolescents with type I and type II of bipolar disorder. *Australasian physical & engineering sciences in medicine*, 2015. 38: p. 551-559.
- [38] Mohammadi, M.R., et al., EEG classification of ADHD and normal children using non-linear features and neural network. *Biomedical Engineering Letters*, 2016. 6: p. 66-73.
- [39] Khaleghi, A., et al., Computational neuroscience approach to psychiatry: A review on theory-driven approaches. *Clinical Psychopharmacology and Neuroscience*, 2022. 20(1): p. 26.
- [40] Afzali, A., et al., Automated major depressive disorder diagnosis using a dual-input deep learning model and image generation from EEG signals. *Waves in Random and Complex Media*, 2023: p. 1-16.
- [41] Khaleghi, A., et al., Abnormalities of alpha activity in frontocentral region of the brain as a biomarker to diagnose adolescents with bipolar disorder. *Clinical EEG and neuroscience*, 2019. 50(5): p. 311-318.
- [42] Xanthopoulos, P., et al., Linear discriminant analysis. *Robust data mining*, 2013: p. 27-33.
- [43] Gragnaniello, D., et al., Local contrast phase descriptor for fingerprint liveness detection. *Pattern Recognition*, 2015. 48(4): p. 1050-1058.
- [44] Jia, X., et al., Multi-scale local binary pattern with filters for spoof fingerprint detection. *Information Sciences*, 2014. 268: p. 91-102.
- [45] Jia, X., et al. Multi-scale block local ternary patterns for fingerprints vitality detection. in *2013 international conference on biometrics (ICB)*. 2013. IEEE.
- [46] Johnson, P. and S. Schuckers. Fingerprint pore characteristics for liveness detection. in *2014 International Conference of the Biometrics Special Interest Group (BIOSIG)*. 2014. IEEE.

# Providing a Hybrid and Symmetric Encryption Solution to Provide Security in Cloud Data Centers

Desong Shen\*

School of Business and Trade, Anhui Wenda University of Information Engineering, Hefei, 230121, China

**Abstract**—One of the most crucial components of information technology infrastructure in the modern world is cloud data centers. Customers have access to these data centers' infrastructure and software, which enable them to store and process massive amounts of data. However, the security and protection of private data in cloud data centers is a serious problem that needs effective and c solutions. Security and privacy issues exist because cloud computing outsources the processing of sensitive data. Consumer worries about cloud infrastructure security remain, particularly those related to data privacy. A thorough analysis of research efforts in the area of cloud security is the main objective of this study. In order to do this, a variety of models were evaluated, their advantages and disadvantages were identified, and a viable security solution based on symmetric algorithms was put forth. The original text in the proposed solution (Hybrid encryption algorithm) is first encrypted using the faster symmetric key method AES, and then its key is encrypted using the faster asymmetric key scheme RSA. This increases efficiency and speed. This method will shorten the time required for data encryption while enhancing its security. The final step was implementing the desired solution in the Eclipse software environment and comparing it against the Blowfish and RSA algorithms. The evaluation's findings indicate that the solution is more advantageous, which has resulted in a nearly two-fold decrease in execution time and a marked increase in throughput when compared to the RSA algorithm. Additionally, the execution time has shrunk, and throughput has been vastly improved compared to the Blowfish method.

**Keywords**—Hybrid encryption algorithm; security; cloud computing; symmetric algorithms

## I. INTRODUCTION

In the ever-evolving IT landscape, cloud data centers stand as giants of innovation, transforming the way we store, process and access data on a scale previously unimaginable. Their ubiquitous presence in our digital lives has ushered in an era of unparalleled convenience, accessibility, and efficiency. However, in the midst of this digital utopia, a looming specter looms – the great challenge of protecting sensitive data in cloud environments. The relentless advancement of technology has brought with it increased security concerns, and data protection has become a critical battleground across the vast expanses of cloud data centers. As we unlock the enormous potential of cloud computing, we are acutely aware of the vulnerabilities it presents. The imperative to fortify data storage fortresses against malicious intrusions and data breaches has never been more prominent. Since the start of its operation, the Internet has seen several changes, some of which have altered how people live today because cloud computing offers consumers a wide range of facilities as a service, this

new technology has swiftly gained popularity [1]. Naturally, any modification and fresh idea in the technological world has its advantages, drawbacks, and issues [2]. This rule applies to using cloud computing as well [3]. We can include the absence of time and location constraints, easy resource sharing, and decreased capital and operating expenses as benefits of cloud computing [4]. Because cloud computing offers scalable resources as a service over the Internet, it has created a number of difficulties for the field's professionals, including data protection or preservation [5]. That is privacy. Early adopters still hesitate to move their businesses to the cloud despite the entire buzz [6]. The difficulties associated with data privacy and information protection continue to disrupt the cloud computing market, and security is one of the primary problems that are slowing down the expansion of cloud computing [7]. Other crucial and major elements of the current model shouldn't be threatened or jeopardized by a new model whose goal is to enhance its features [8]. Cloud architecture poses risks to the security of these technologies when they are employed in a cloud environment [9]. Users of cloud services should be cautious and knowledgeable about the risks associated with data breaches in this novel environment [10]. User data is typically safeguarded from hackers using a variety of encryption techniques in order to assure security and secrecy. This technique has also been applied in cloud computing settings. However, reliability is not ensured when numerous services are used concurrently to carry out operations like combining functions [11]. For instance, if a malicious application interferes with the service used by other customers then the numerous cloud platforms are shared by many clients. Other people's environments are also impacted by it [12]. Common security threats for cloud computing include integrity, availability, and confidentiality [13]. Attacks in this system can be split into two categories, assaults from both internal and external parties. Internal assaults are those in which the perpetrator seeks to get access to the network and its operations by obtaining virtual machine control, knowing the password or other authentication credentials, or both [14].

In contrast, attackers that use external attacks want to distribute false routing information or stop nodes from offering services. An internal intruder poses a greater threat than an outside one. A secure link between data centers and users is created concurrently using symmetric encryption as well as hybrid encryption [15]. With the aid of this technique, key structures for data encryption and decryption in cloud environments are made to be secure and dependable [16]. The benefits of utilizing hybrid and symmetric encryption in cloud data centers include boosting user privacy protection, lowering the risk of cyberattacks, strengthening protection against

intrusion, and increasing security and protection of sensitive information retrieval. Security mechanisms against hackers (attackers) including: Firewalls and intrusion detection systems (IDS), access control and authentication, encryption, security patch management, intrusion response plan, network segmentation, security information and event management (SIEM), and monitoring and logging.

Additionally, the usage of symmetric and hybrid encryption boosts the effectiveness of cloud data centers [17]. Additionally, symmetric and hybrid encryption can withstand physical assaults. Information is kept in decryption mode, and the likelihood of unwanted access to them is decreased by utilizing detection and prevention techniques. These techniques enable cloud data centers to deliver services more dependably while limiting illegal data access [18]. When using symmetric encryption algorithms such as AES, a separate MAC algorithm or an authenticated encryption mode (such as AES-GCM or AES-CCM) is used to simultaneously ensure data confidentiality and integrity/authentication. These modes incorporate MAC functions in the encryption process to achieve data integrity protection. Also, when a secure cryptographic system is implemented, a well-established MAC algorithm (such as HMAC, CMAC) or an authenticated encryption mode should be considered to protect data integrity and validity in addition to encryption. In order to guarantee security and boost efficiency in cloud data centers, a novel approach utilizing hybrid and symmetric encryption is given in this article. Combining many distinct encryption techniques at once is known as hybrid encryption. By using this technique, it is possible to increase security and defend against decoding attempts. A secure link between data centers and users is created concurrently using symmetric encryption as well as hybrid encryption. With the aid of this technique, key structures for data encryption and decryption in cloud environments are made to be secure and dependable. The author's contribution to this work can be summed up as follows:

- Offering an integrated solution based on the RSA and AES algorithms to improve security.
- Shortening the duration of encryption operations.

The remainder of the essay is structured as follows: A summary of earlier efforts is provided in the Section II. Section III goes into further detail about the suggested approach. The evaluation and effectiveness of the suggested method are described in the Section IV. Section V contains the conclusion and recommendations for further research.

## II. RELATED WORKS

One of the most talked-about subjects in the IT world is cloud computing. Due to the cloud's vast resources and low entry barrier, it is being enthusiastically embraced by many new businesses. The cloud, however, has both pros and cons, just like any other topic. Information about cloud users is saved remotely. Therefore, one of the primary concerns of any firm contemplating a move to the cloud is cloud data security. Firewalls and VPNs (virtual private networks) are two of the most popular ways for data owners to protect their information at home or the office. The company is the data owner, but it

uses uncontrolled cloud servers to store sensitive information, and its users can access this information when needed. For this reason, there is a security risk associated with storing client data elsewhere.

As a result, safeguarding data in the cloud has emerged as a key field of study [19]. Various cryptographic methods, including AES (a symmetric cryptographic method), SHA-1 (a hashing method), and ECC, are used in [20], with data first being organized into categories according to its sensitivity and significance. In the asymmetric cryptography method of elliptic curve cryptography, most existing works rely on a single key for encryption and decryption, making them vulnerable to a wide range of well-known harmful attacks. As a result, the hybrid algorithm we developed uses two independent keys applicable to any encoding/decoding process. If you wish to access data stored in the cloud, you'll need to sign up with both your cloud service provider and the cloud's owner which is required for decrypting cloud-based information. The purpose of this research [21] was to create a novel approach to cloud data security based on hybrid encryption architecture. Encrypting user data before it is sent to the cloud is an active study area because of the prevalence of malicious actors in this environment.

Since the scope of cloud computing security concerns extends to data access control, identity management, auditing, integrity control, and risk management, a hybrid cryptosystem was developed to address these concerns. Data privacy is addressed with the symmetric Blowfish algorithm rather than the asymmetric RSA scheme. Authentication is the focus of the algorithm. The Secure Hash-2 algorithm is also used in this method; therefore, the data may be trusted. Based on the current research results, it has been determined that the suggested method offers both high security for data transmission over the Internet and easy, on-demand network access to the manufacturer's shared pool of computer resources, including the latter two in particular. Due to concerns over data security, many large companies are hesitant to adopt cloud computing services. There have been numerous incidents of cloud security breaches documented over the past few years, despite the cloud service provider's claim of having a solid third-party security system. Therefore, for cloud service providers to succeed, they must have stringent safety measures. Strong security for user data is proposed in [22] via a two-layer agent-based hybrid framework that combines symmetric and asymmetric key methods. The risk of data misuse by the cloud service provider is also removed because the user alone controls the decryption process. This framework improves security without slowing down the virtual machine's processing speed since the two cryptographic algorithms utilized are optimized for low key size, low encryption time, and high speed. In [23], we examine hybrid cryptography from 2014 until the beginning of 2019. Eight are based on a tabular survey format that is easy to use, while the remaining 12 are comprehensive polls. The primary goal of this review paper is to expand the knowledge base of novice cryptography researchers, students, and practitioners. The lack of attention paid to user authentication and the improper usage of hybrid algorithms is the area where more study is needed. As a result of all this, cloud customers are beginning to worry about the

security of their data while it is being stored on these outside managed servers. Information security is necessary to prevent these data breaches and other risks. Information security relies heavily on encryption. The user employs a couple of different encryption algorithms to keep their cloud data safe. Researchers look at trust and its application in distributed computing [24]. A summary of proposed trust models for various distributed system types is provided. The proposed trust management systems for cloud computing are examined with a focus on their capabilities, realistic heterogeneous cloud applicability, and implementation viability. In actuality, data security refers to the safeguarding of information's availability, completeness, and secrecy.

Research [25] ensures that, when necessary (Availability), only authorized users have access to accurate and comprehensive information (Completeness). Information security aims to shield data and information systems from misuse, failure, disclosure, and unauthorized access. Based on the Security as a Service (SECaaS) concept, they suggested multi-layer and multi-level encryption architecture for cloud computing [26]. This article also makes the point that the implementation of this architecture is adaptable to scalable systems with various needs and can unite heterogeneous networks and different operating systems. K-anonymity, an encryption model to safeguard personal information, is provided in [27]. By employing this technique, user data can be protected from unauthorized disclosure. The term "data" in this article refers to specific, unique information that is conceptually arranged as a table with rows for reports and columns for strings. An outline of the cloud's security issues and major difficulties can be found in [28]. By the end of this piece, they have determined that a significant portion of data security may be achieved by the employment of cryptographic methods.

Additionally, it has been found that using both symmetric and asymmetric encryption at once can speed up message transmission and identity verification considerably. The employment of extensible identity recognition protocol in cloud computing is plagued by issues listed in [29]. Cloud computing authentication issues have been resolved using EAP. This technique nullifies dangers from data manipulation, DoS assaults, and identity theft. However, in this manner, a powerful algorithm and encryption are required for the cloud environment, ensuring that the client's data and the data transferred between the client and the cloud provider are encrypted. The EAP approach also has additional issues. A framework for the authentication choice was supplied by the study conducted in [30], referred to as the "Trust Cube" in this study. A high-level framework of authentication procedures is offered in this solution. The client device, the data collector, the authentication engine, and the authentication consumers are taken into account in this architecture. Each of these participants must be validated before data can be sent through the authentication engine.

### III. SUGGESTED METHOD

The suggested approach is broken up into two stages. The first stage focuses on the transfer and secure storage of data on the cloud. Data is only made available to the authorized user after passing all security mechanisms in the second step, which handles data retrieval from the cloud and data validation and integrity. A hybrid encryption technique is employed in both phases as the best option for data encryption and decryption, which is done by using the data from the user side to the server or vice versa. The proposed model and method to provide a general framework to guarantee data security and comprehensiveness are shown in the general structure in Fig. 1. You can see the overall layout and the different types of actions done on the data at each level in this structure.

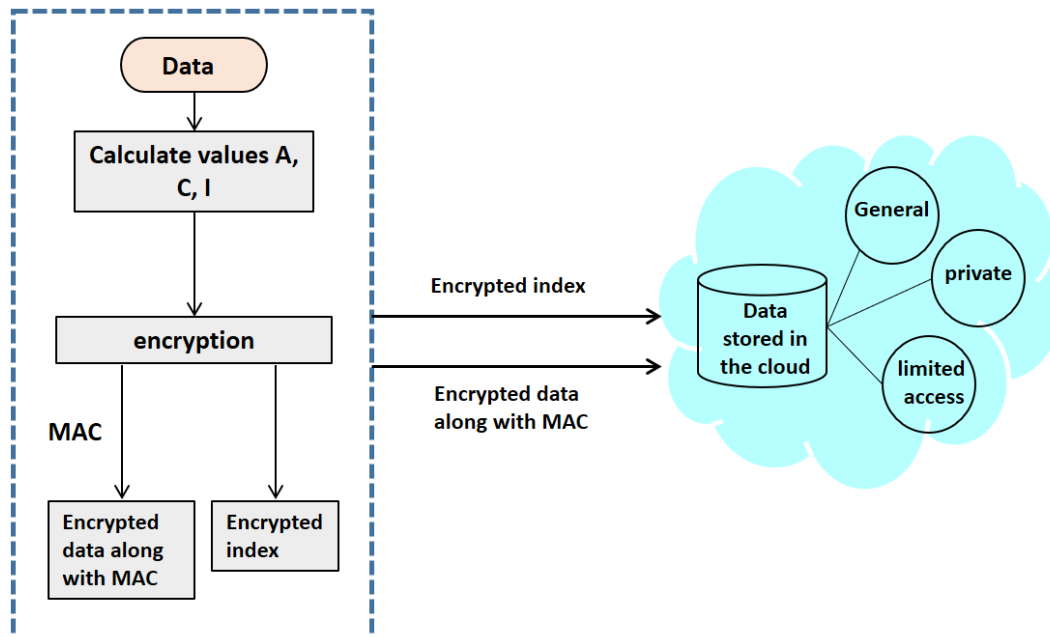


Fig. 1. General outline of the proposed model.

In actuality, this encryption method combines the strengths of AES and RSA, two symmetric encryption methods, to establish the key cryptographic parameters. There was good assurance of confidentiality, integrity, and availability. A systematic and comprehensive method was employed to explore the complexities of cloud data center security and evaluate the effectiveness of our proposed hybrid encryption solution. This method is designed to increase clarity and transparency and allow replication and validation of our research findings. The key components of our research method are as follows: Problem identification, model design, implementation, testing, security analysis, comparison with existing methods, data analysis and presentation of results.

In applying this systematic methodology, our aim was to ensure the robustness and validity of our research. The step-by-step approach brought clarity to our research process and facilitated the evaluation of the effectiveness of our proposed hybrid encryption solution in increasing security and efficiency in cloud data centers.

A. Combination of two encryption algorithms

Fig. 2 displays the overall architecture of the suggested method for integrating two techniques employing RSA asymmetric and AES symmetric encryption.

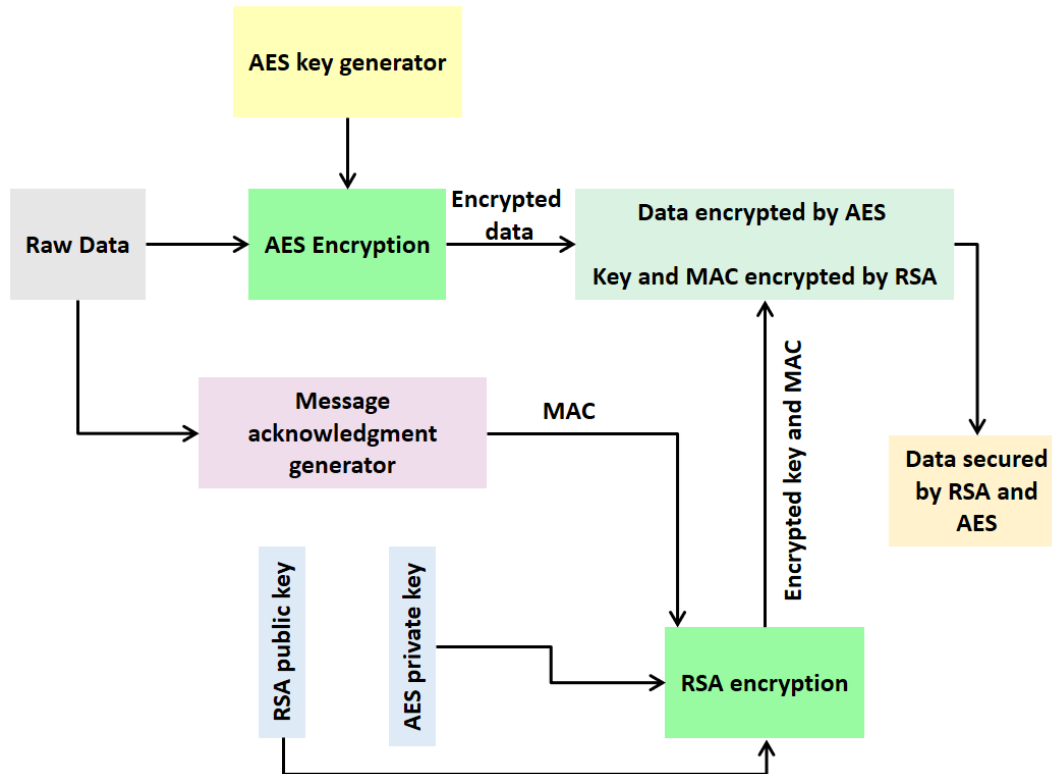


Fig. 2. The general framework of the combination of two cryptographic algorithms.

Proposed hybrid algorithm
Input: Data [] array of n integers storing data for the protection section. Where Data is an integer array containing the nonnegative integers A, C, I, AES and RSA. Output: Information sorted by category for the relevant section. For a to x C [a] = value of confidentiality. I [a]= value of integrity A [a]= value of availability Calculate AES and RSA [i] = (s [a] + (1/v[a])*10+I [a] )/2 For b= 1 to 10 For a= 1 to x If AES and RSA [a] == 1, 2, 3 then S [a] = 3 If AES and RSA [a] == 4, 5, 6 then S [i] = 2 If AES and RSA [a] == 8, 9, 10 then S [a] = 1

Since cloud storage is preferred, methods are provided for storing various types of data there (public, private, restricted access) according to three cryptographic parameters: privacy, accessibility, and security. The value of "C" (Confidentiality) is determined by how much privacy is needed at each data processing step. In contrast, the value of "I" (Integrity) is determined by how well the data is accurate, reliable, and protected against unauthorized changes. Fig. 1's A (Availability) represents how readily available the data needs to be in response to a user's request.

Users of the aforementioned algorithm are tasked with categorizing information according to the three cryptographic parameters C, I, and A. The user is responsible for providing the values for C, I, and A, where  $D_{ata} []$  is the data. Integrity is also directly tied to security and secrecy, while security is inversely related to availability. This "SR" number determines which of Fig 2's three sections the information belongs in S3 [Public], S2 [Private], or S1 [Owner's Limited Access].

In this solution, the interaction between the user and the cloud servers is considered the main step in which the user must be registered as a cloud client. If he is a registered user, the password check process is performed. Otherwise, you must first register using the service provider, and after confirming the user's authentication process, his public key will be sent to the server to encrypt the data. The scheduling unit of runtime coordination can be categorized as follows from the standpoint of runtime scheduling granularity:

- The complete application framework, which consists of all the entities that are used for execution and collaboration as well as the required external containers. Take a Hadoop, for instance.
- An instance of an application workload, comprising all the entities involved in distributed execution, is referred to as an application instance. Consider a Hadoop job.
- A single executor, which is typically a local OS process, is the internal execution entity of an application instance. Take a Hadoop task, for instance.

#### IV. EVALUATION RESULTS

These tests assess the effectiveness of data encryption at various sizes. Fig. 3 to 11 and Tables I to III exhibit the experiment's outcomes. The differences between the implementation of cryptographic operations in the combined mode and in comparison, with other methods may be readily recognized through the evaluation of the graphs. The comparison of the combined mode's execution times clearly demonstrates improved efficiency and optimal execution. In reality, the encryption procedure has been completed in half the time required by the RSA technique, making the execution time more than two times faster. The high execution time of asymmetric coding techniques is actually one of their key issues, which has a significant impact on performance. Although the throughput is somewhat less optimized when compared to the Blowfish method, the execution time is almost 30 milliseconds less, demonstrating that the combined solution is more optimal.

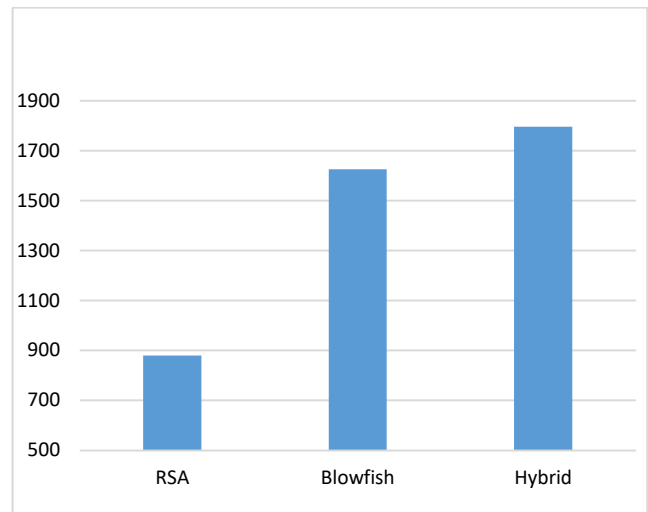


Fig. 3. Throughput - kilobytes per second (data size: 512 kilobytes).

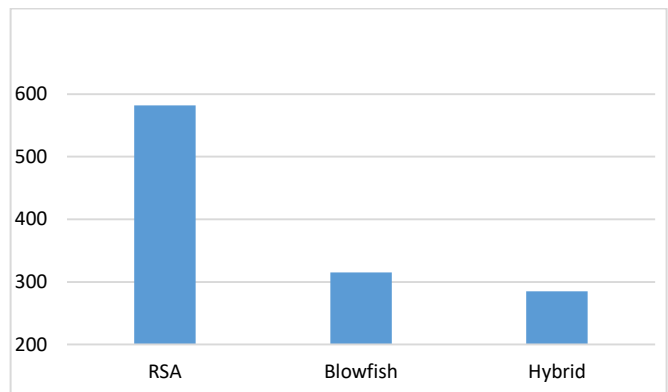


Fig. 4. Execution time - milliseconds (data size: 512 KB).

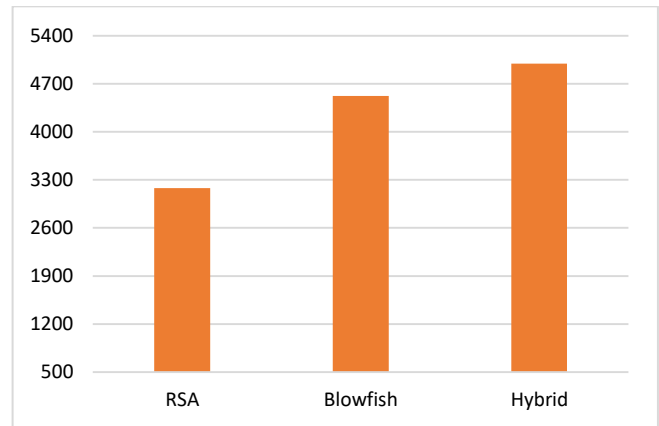


Fig. 5. Throughput - kilobytes per second (data size: 2048 kilobytes).

TABLE I. EXECUTION TIME AND THROUGHPUT (DATA SIZE: 512 KB)

	Hybrid	Blowfish	RSA
Throughput	1796	1625	880
Execution time (milliseconds)	285	315	582

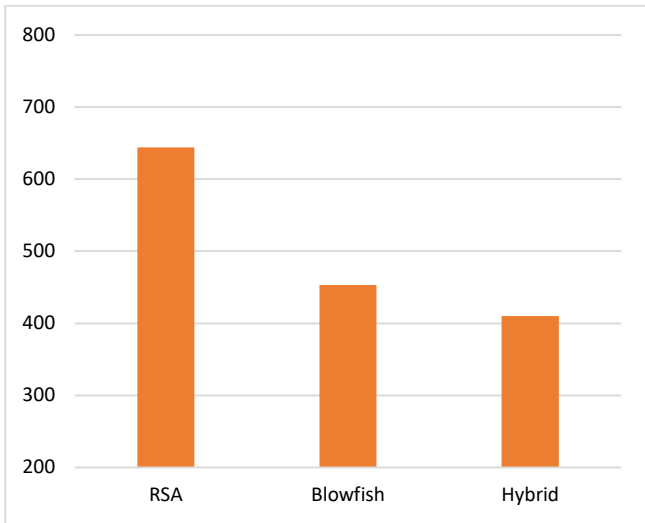


Fig. 6. Execution time - milliseconds (data size: 2048 KB).

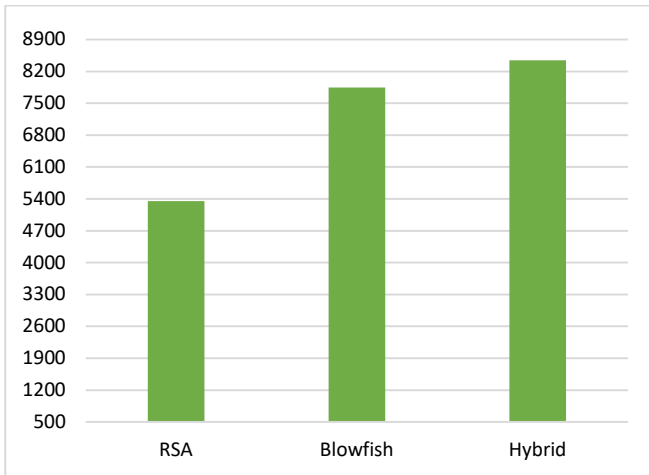


Fig. 7. Throughput - kilobytes per second (data size: 4096 kilobytes).

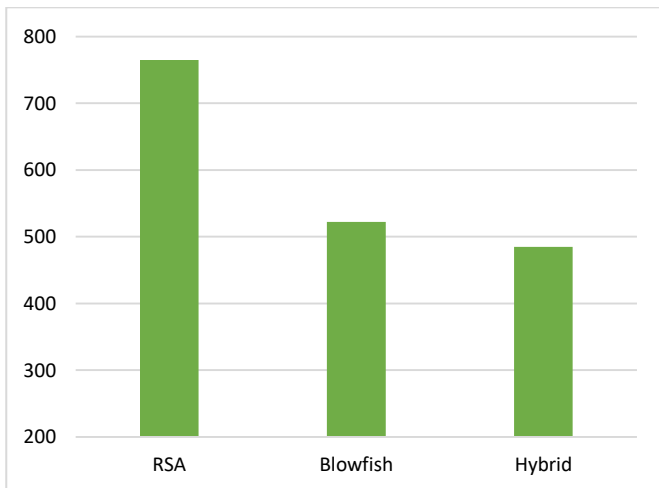


Fig. 8. Execution time - milliseconds (data size: 4096 KB).

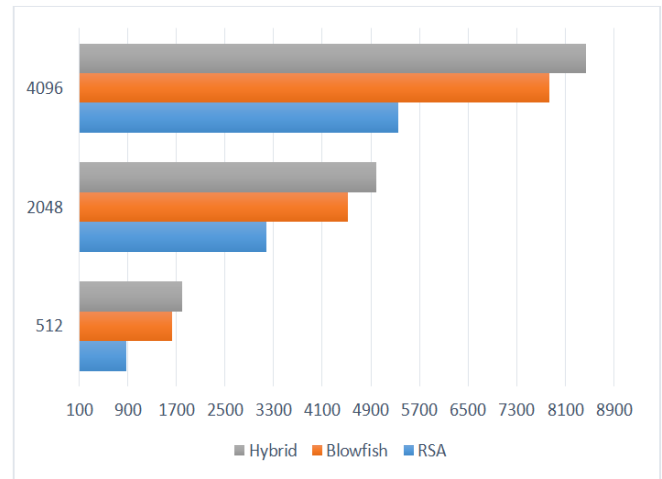


Fig. 9. Throughput rate in different data sizes - kilobytes per second.

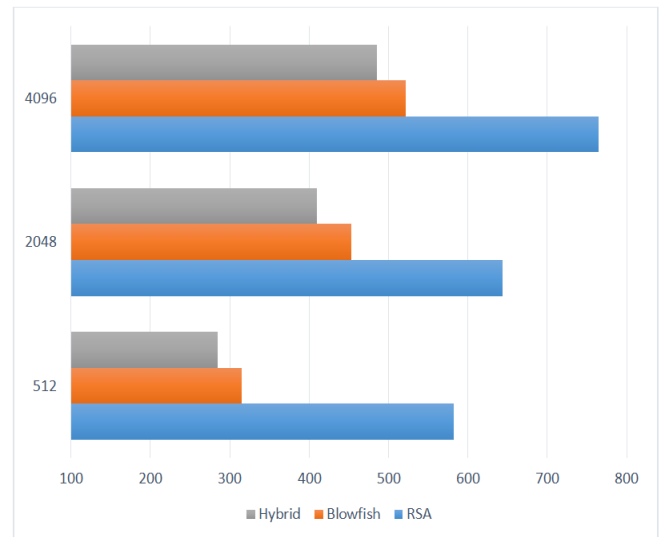


Fig. 10. Execution time in different data sizes - in milliseconds.

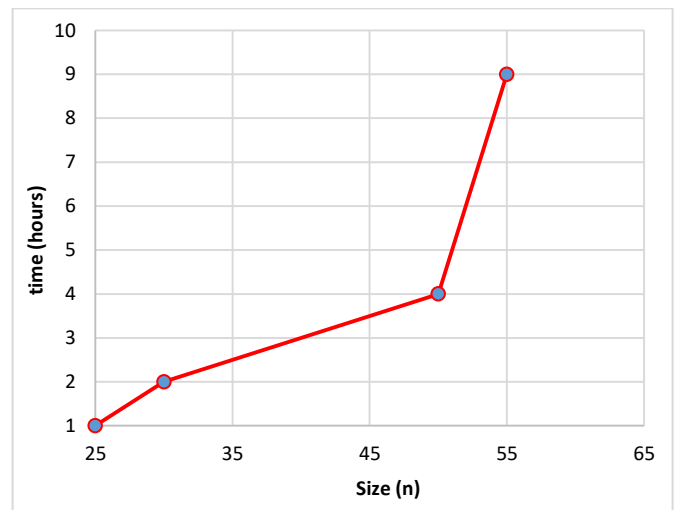


Fig. 11. The time required to decompose n into prime factors.

Due to the fact that the size of the data that can be encrypted can affect the time of the operation, in the second test, the size of the data has been increased to two megabytes (2048 kilobytes) to compare the efficiency of algorithms in this volume of data. As can be seen from Table II, the proposed solution performs much better than the RSA algorithm. In addition to the better throughput, the execution time is reduced by about 214 milliseconds. Compared to the Blowfish algorithm, the execution time has decreased by about 43 milliseconds and at the same time, the transmittance has also increased. This shows that the proposed solution works faster than these two solutions.

TABLE II. RUNNING TIME AND THROUGHPUT (DATA SIZE: 2048 KB)

	Hybrid	Blowfish	RSA
Throughput	4995	4521	3180
Execution time (milliseconds)	410	453	644

TABLE III. EXECUTION TIME AND THROUGHPUT (DATA SIZE: 4096 KB)

	Hybrid	Blowfish	RSA
Throughput	8445	7847	5354
Execution time (milliseconds)	485	522	765

Charts 9 and 10 clearly demonstrate that the proposed solution is more optimal in the last test, which compared the desired algorithms with a data amount of 4096 kilobytes. The proposed solution was able to perform the operation in less time than the other two algorithms, as shown in Table III. As the amount of data increases, the throughput likewise increases, but it impacts the encryption time and results in more time required.

The RSA technique is asymmetric, so it naturally takes a lot of time to produce the public and private keys and perform the encryption process, so its execution time is significantly higher. However, the other solution takes less time because of its symmetry. The results also demonstrate how little implementation there is.

Finally, in Fig. 9 and 10, a general comparison between execution times and throughput in different data sizes can be seen.

As can be seen, the processing action requires greater time for the initial execution than for successive executions, which indicates the necessity for warming up or preparation time. This time will be shorter for the application of larger data, yielding the desired outcome. The block size of processable threads was set at 256 for the processing resources that will be accessible for cryptographic operations, albeit this number may change depending on the resources' makeup. The greatest quantity of data that can be processed at any given time for encryption and decryption operations will be equivalent to 256 gigabytes. On the other hand, each processable thread can hold 64 bits of data (8 bytes).

### A. Time of Failure

This section evaluates the failure time of the suggested solution because, in addition to boosting efficiency, increasing security is one of the key goals of this research. Failure time is the amount of time needed to locate the algorithm's secret key and subsequently decrypt it. It is obvious that the answer is more secure the longer this period is. Although all encryption techniques can indeed be defeated, what matters is how long the information should be decrypted and what tools should be used. In relation to the suggested hybrid approach, since the RSA technique encrypts the AES encryption key, if the RSA can be decrypted, the AES key may also be retrieved, allowing for the discovery of the original content.

It should be noted that there are only a few ways to decrypt text using the RSA algorithm, the primary one being the breakdown of  $n$  into prime factors. In this instance,  $n$  and  $e$  are likewise presumed to be provided together with the RSA public key. Decomposing  $n$  into its prime factors,  $p$  and  $q$ , is the initial step at this point. Actually, the main and most challenging element of decrypting an RSA key is this step. Mathematical computations have demonstrated that it would take more than seven months to find the prime factors of a number with 155 digits, for instance, even using the fastest computers. The crucial point is that by choosing a larger key (choosing larger  $p$  and  $q$  numbers) when employing RSA, the work of analyzing  $n$  can be made much more challenging for new computers, regardless of how fast and adept they are at handling huge numbers. The outcomes of the decomposition of various values of  $n$  into prime components, along with the time needed to complete it, are presented.

According to evaluations, if the number of digits is taken to be equal to 200, four million years of time, as can be seen in the above figure, the time of decomposition might increase exponentially the greater the value of  $n$ . The ability to decode the data is required. Fig. 11 makes it evident that picking larger digits can actually make the analysis time; as a result, the RSA decryption is unfeasible in a short amount of time.

### B. Efficiency Comparison

An effective cloud data security model will solve all cloud computing's potential issues, allowing its advantages to soar to new heights while shielding its owner's data from as many threats as feasible. In Table IV, we can see how the suggested model stacks up against competing approaches to data protection.

After the security parameters for the contrasted approaches were implemented, Fig. 12 depicts the level of security. As can be seen, the security factor of the suggested method is higher compared to previous ways due to the rise in the volume of data in the horizontal axis, as well as the usage of data classification and encryption technology.



TABLE IV. FACTORS OF THE PROPOSED METHOD COMPARED TO SIMILAR WORKS

Factors	[3]	[31]	[9]	[13]	[21]	[24]	[25]	[29]	This work
Authentication of Storage Providers	n	y	y	y	n	n	y	n	y
Confidentiality	n	y	n	y	y	n	y	n	y
Non-repudiation	n	n		y	y	y	n	n	y
Safe even if the user's credentials are compromised	y	n	y	n	n	y	n	y	y
Authorization	y	y	y	n	n	y	y	y	y
Encryption	y	n	y	y	y	y	n	n	y
Identifying and verifying	y	y	n	n	n	n	y	y	y
Integrity	n	n	y	n	n	n	y	n	y
File indexing	y	y	n	y	y	n	n	y	y
Lookup by Keywords	n	n	y	n	y	y	n	y	y

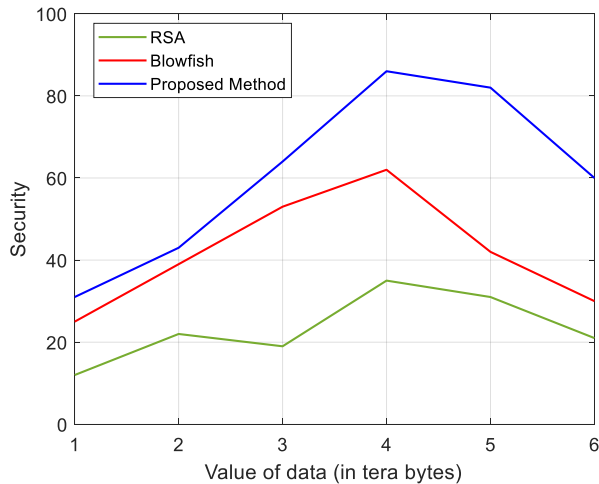


Fig. 12. Comparison of methods for security assessments.

This research includes a hybrid approach using both RSA and AES algorithms for encryption in cloud data centers, which is summarized as follows:

1) *Execution time and throughput analysis:* 512 KB data size: The proposed hybrid solution performs significantly better than the RSA algorithm when encrypting 512 KB data. In particular, it notes that the execution time is cut in half compared to RSA, which represents a significant improvement in speed. Furthermore, although the throughput is slightly lower than the Blowfish method, it is noted that the execution time is approximately 30 milliseconds shorter, indicating that the hybrid solution offers a more optimal balance between speed and security.

2048 KB data size: When the data size increases to 2048 KB (2 MB), the hybrid solution still outperforms RSA by a large margin. The execution time is reduced by approximately 214 ms compared to RSA, indicating that the hybrid approach is significantly faster. The article also mentions improvements in throughput.

4096 KB data size: In the final evaluation with a data size of 4096 KB (4 MB), the hybrid solution still shows its superiority over RSA and Blowfish. Runtimes are reported to be approximately 280ms faster than RSA and 37ms faster than Blowfish, showing consistent and significant speed benefits.

2) *Breakdown time analysis:* In this research, an analysis of the RSA algorithm's failure time is presented, which is a measure of how long it takes to decrypt RSA-encrypted data. This emphasizes the importance of choosing large key sizes to make computational decryption impossible in a reasonable amount of time. The presented results show that with a large enough key size (e.g., 200 digits), the decryption time can increase to millions of years, which highlights the security advantage of using larger key sizes in RSA encryption.

3) *Efficiency comparison:* In the following, the proposed hybrid model has been compared with other methods in terms of security evaluation. It uses a security factor as a benchmark and shows that the hybrid model consistently shows higher security levels compared to alternatives. This suggests that the hybrid approach provides a superior balance between security and efficiency.

The proposed method points out that the hybrid approach combines the strengths of AES and RSA for cryptographic parameters, which results in ensuring confidentiality, integrity, and availability. While providing runtime and throughput data, it is helpful to include specific numerical results, charts, or graphs to visually demonstrate these performance improvements.

## V. CONCLUSION

In addition to ensuring security, this article's major objective is to make encryption and decryption processes more effective. As a result, a hybrid approach based on the RSA and AES algorithms was presented to speed up encryption procedures while boosting security. As a result, the message confirmation code and the RSA method are used to secure the key after the main data has been encrypted using the AES technique. With this technique, the potential for decreasing the time required for the encryption procedure is also well established, in addition to improving security. Finally, a hybrid algorithm was constructed and tested using the Java programming language and the Eclipse programming environment, and all tests conducted on the provided solution show that it is superior to alternative approaches; there were strategies. Therefore, the encryption procedure was completed in half the time required by the RSA technique in the first test with a data volume of 512 kilobytes. Although the throughput was higher than the Blowfish method, the execution time was almost 30 milliseconds lower, demonstrating the combined solution's superiority. The proposed approach outperformed the

RSA technique in the second trial, which used a data volume of 2048 KB. Compared to the Blowfish algorithm, throughput increased more than two times, and execution time dropped by more than 214 milliseconds.

Additionally, the throughput is higher, and the execution time is decreased by 43 milliseconds, showing that the proposed method operates more quickly than the two alternatives. The proposed solution outperformed the other two algorithms in the final evaluation with a data volume of 4096 kilobytes, resulting in execution times that were about 280 milliseconds faster than those of the RSA algorithm and 37 milliseconds faster than those of the Blowfish algorithm, respectively. In comparison to these three methods, it demonstrates total optimality. The failure time of the RSA method, which is in charge of protecting the private key, was also examined to assess the solution's level of security. The findings of the experiment demonstrate that by using large keys, it can be decoded in a fair amount of time.

It is feasible to add to the described algorithm for future research and to broaden the suggested solution; it also makes use of a trusted third-party (TTP) based model because TTPs are frequently used in both commercial and cryptographic digital transactions, particularly in those involving a CA that gives a digital identity certificate to one of the two parties. Accordingly, the user is first verified using TTP-based protocols and then, after receiving the authentication certificate, can access the system and view and decode the encrypted data. At that point, the CA becomes a trusted third party for issuing certificates.

#### ACKNOWLEDGMENT

Anhui Province Scientific research project "Research on the Innovative Development Path of Anhui Cross-border E-commerce under the Background of Free Trade Zone Construction" Project No.: SK2021A0814.

#### REFERENCES

- [1] M. Kaur and R. Aron, "Energy-aware load balancing in fog cloud computing," *Mater Today Proc*, 2020.
- [2] O. Y. Abdulhammed, "Load balancing of IoT tasks in the cloud computing by using sparrow search algorithm," *J Supercomput*, vol. 78, no. 3, pp. 3266–3287, 2022.
- [3] M. M. S. Maswood, M. D. R. Rahman, A. G. Alharbi, and D. Medhi, "A novel strategy to achieve bandwidth cost reduction and load balancing in a cooperative three-layer fog-cloud computing environment," *IEEE Access*, vol. 8, pp. 113737–113750, 2020.
- [4] M. Trik, S. P. Mozaffari, and A. M. Bidgoli, "Providing an adaptive routing along with a hybrid selection strategy to increase efficiency in NoC-based neuromorphic systems," *Comput Intell Neurosci*, vol. 2021, 2021.
- [5] Fabian 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
- [6] A. Celesti, M. Fazio, A. Galletta, L. Carnevale, J. Wan, and M. Villari, "An approach for the secure management of hybrid cloud-edge environments," *Future Generation Computer Systems*, vol. 90, pp. 1–19, 2019.
- [7] Ning Xu, Zhongyu Chen, Ben Niu, and Xudong Zhao. Event-Triggered Distributed Consensus Tracking for Nonlinear Multi-Agent Systems: A Minimal Approximation Approach, *IEEE Journal on Emerging and*

- Selected Topics in Circuits and Systems, DOI: 10.1109/JETCAS.2023.3277544, 2023.
- [8] M. K. Neha, "Enhanced security using hybrid encryption algorithm," *International Journal of Innovative Research in Computer and Communication Engineering*, vol. 4, no. 7, pp. 13001–13007, 2016.
- [9] 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*, 360(14), 10728-10744.
- [10] P. Crocker and P. Querido, "Two factor encryption in cloud storage providers using hardware tokens," in 2015 IEEE Globecom Workshops (GC Wkshps), IEEE, 2015, pp. 1–6.
- [11] 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.
- [12] B. Seth, S. Dalal, and R. Kumar, "Hybrid homomorphic encryption scheme for secure cloud data storage," *Recent Advances in Computational Intelligence*, pp. 71–92, 2019.
- [13] Wenjing Wu, Ning Xu, Ben Niu, Xudong Zhao and Adil M. Ahmad, Low-Computation Adaptive Saturated Self-Triggered Tracking Control of Uncertain Networked Systems, *Electronics*, 12(13), 2771, 2023.
- [14] M. Samiei, A. Hassani, S. Sarspy, I. E. Komari, M. Trik, and F. Hassanpour, "Classification of skin cancer stages using a AHP fuzzy technique within the context of big data healthcare," *J Cancer Res Clin Oncol*, pp. 1–15, 2023.
- [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] J. Sun, Y. Zhang, and M. Trik, "PBPHS: a profile-based predictive handover strategy for 5G networks," *Cybern Syst*, pp. 1–22, 2022.
- [17] M. Trik, H. Akhavan, A. M. Bidgoli, A. M. N. G. Molk, H. Vashani, and S. P. Mozaffari, "A new adaptive selection strategy for reducing latency in networks on chip," *Integration*, vol. 89, pp. 9–24, 2023.
- [18] Haoyan Zhang, Xudong Zhao, Huangqing Wang, Ben Niu, Ning Xu, Adaptive Tracking Control for Output-Constrained Switched MIMO Pure-Feedback Nonlinear Systems with Input Saturation, *Journal of systems science & complexity*, 36: 960–984, 2023.
- [19] Abouzarkhanifard, A., Chimeh, H. E., Al Janaideh, M., & Zhang, L. (2023). Fem-inclusive transfer learning for bistable piezoelectric mems energy harvester design. *IEEE Sensors Journal*, 23(4), 3521-3531.
- [20] M. Trik, A. M. N. G. Molk, F. Ghasemi, and P. Pouryeganeh, "A hybrid selection strategy based on traffic analysis for improving performance in networks on chip," *J Sens*, vol. 2022, 2022.
- [21] Wang, Z., Jin, Z., Yang, Z., Zhao, W., & Trik, M. (2023). Increasing efficiency for routing in Internet of Things using Binary Gray Wolf Optimization and fuzzy logic. *Journal of King Saud University-Computer and Information Sciences*, 101732.
- [22] H. Abroshan, "A hybrid encryption solution to improve cloud computing security using symmetric and asymmetric cryptography algorithms," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, pp. 31–37, 2021.
- [23] C. Kota, "Secure File Storage in Cloud Using Hybrid Cryptography," Available at SSRN 4209511, 2022.
- [24] V. Goyal and C. Kant, "An effective hybrid encryption algorithm for ensuring cloud data security," in *Big Data Analytics: Proceedings of CSI 2015*, Springer, 2018, pp. 195–210.
- [25] A. Kumar, V. Jain, and A. Yadav, "A new approach for security in cloud data storage for IOT applications using hybrid cryptography technique," in 2020 international conference on power electronics & IoT applications in renewable energy and its control (PARC), IEEE, 2020, pp. 514–517.
- [26] 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.
- [27] P. B. Regade, A. A. Patil, S. S. Koli, R. B. Gokavi, and M. S. Bhandigare, "SURVEY ON SECURE FILE STORAGE ON CLOUD USING HYBRID CRYPTOGRAPHY," *International Research Journal of*

- Modernization in Engineering Technology and Science, vol. 4, no. 06, 2022.
- [28] S. Gokulraj, P. Ananthi, R. Baby, and E. Janani, "Secure File Storage Using Hybrid Cryptography," Available at SSRN 3802668, 2021.
- [29] S. Rehman, N. Talat Bajwa, M. A. Shah, A. O. Aseeri, and A. Anjum, "Hybrid AES-ECC model for the security of data over cloud storage," *Electronics (Basel)*, vol. 10, no. 21, p. 2673, 2021.
- [30] J. Lei, Q. Wu, and J. Xu, "Privacy and security-aware workflow scheduling in a hybrid cloud," *Future Generation Computer Systems*, vol. 131, pp. 269–278, 2022.
- [31] L. Huang, K. Feng, and C. Xie, "A practical hybrid quantum-safe cryptographic scheme between data centers," in *Emerging Imaging and Sensing Technologies for Security and Defence V*; and *Advanced Manufacturing Technologies for Micro-and Nanosystems in Security and Defence III*, SPIE, 2020, pp. 30–35.

# A Single-Stage Deep Learning-based Approach for Real-Time License Plate Recognition in Smart Parking System

Lina YU<sup>1\*</sup>, Shaokun LIU

Hebei College of Industry and Technology, Hebei Shijiazhuang, 050091, China

**Abstract**—License plate recognition in smart parking systems plays a crucial role in enhancing parking management efficiency and security. Traditional methods and deep learning-based approaches have been explored for license plate recognition. Deep learning methods have gained prominence due to their ability to extract meaningful features and achieve high accuracy rates. However, existing deep learning-based fire detection methods face challenges in terms of accuracy, real-time requirement, and computation cost, as evident from previous studies. To address these challenges, we propose a single-stage deep learning approach using YOLO (You Only Look Once) algorithm. Our method involves generating a custom dataset and conducting training, validation, and testing processes to train the YOLO-based model. Experimental results and performance evaluations demonstrate that our proposed method achieves high accuracy rates and satisfies real-time requirements, validating its effectiveness for license plate recognition in smart parking systems.

**Keywords**—Smart parking; license plate recognition; deep learning; single-stage detector; Yolo

## I. INTRODUCTION

Smart parking systems have emerged as a transformative solution to address the challenges of urban parking management, aiming to optimize parking resource utilization and enhance the overall parking experience for drivers [1, 2]. These systems leverage advanced technologies to provide real-time parking information, streamline parking processes, and contribute to efficient traffic management in smart cities [3].

License plate recognition (LPR) technology plays a pivotal role in smart parking systems by enabling automated vehicle identification, entry/exit control, and payment processes [4, 5]. LPR involves the detection, extraction, and recognition of license plate information from images or videos. By accurately capturing and processing license plate data, smart parking systems can provide seamless and convenient parking experiences to drivers, optimize resource allocation, and enhance operational efficiency.

Current technologies and recent advances in license plate recognition have significantly enhanced the capabilities of smart parking systems. Various approaches, including computer vision, machine learning, and deep learning, have been employed to improve the performance of LPR systems [6]. Among these technologies, vision-based methods have garnered significant attention from researchers due to their

ability to handle variations in lighting conditions, vehicle types, and license plate designs.

Vision-based methods utilize computer vision techniques to extract and analyze license plate information from images or videos [7]. These methods have shown promising results in terms of accuracy and reliability, making them suitable for real-world deployment in smart cities. The significance of research in this area lies in developing more accurate and efficient LPR systems that can handle complex real-world scenarios and meet the real-time requirements of smart parking systems [8]. On the other hand, deep learning-based methods have emerged as state-of-the-art approaches for license plate recognition in smart parking systems. Deep learning models, particularly those utilizing convolutional neural networks (CNNs), have shown remarkable advancements in various computer vision tasks [9, 10]. However, there are still limitations and challenges in existing deep learning-based approaches for license plate recognition, particularly in achieving real-time processing and high accuracy rates.

To tackle these limitations, further research is needed to investigate alternative approaches that can achieve high accuracy rates in real-time. One such approach is the utilization of single-stage deep learning models, such as the YOLO (You Only Look Once) algorithm [11]. Single-stage models eliminate the need for time-consuming region proposal networks and achieve efficient end-to-end license plate detection and recognition [12]. By exploring single-stage and YOLO-based methods, the research aims to address the current limitations and challenges in deep learning-based license plate recognition for smart parking systems.

In this study, we propose a YOLO-based method with a custom dataset generation to tackle the identified research gap and meet the real-time requirements of license plate recognition in smart parking systems. The YOLO-based model is generated using a custom dataset, encompassing various challenging scenarios encountered in smart parking environments. The model is trained, validated, and tested to evaluate its performance in terms of accuracy, speed, and robustness.

The research contributions of this study lie in identifying the research gap in deep learning-based license plate recognition for smart parking systems and proposing a YOLO-based method to address this gap. Additionally, comprehensive experimental evaluations are conducted to validate the proposed method, assessing its performance in real-world

scenarios. The contributions of this research are significant for the development of accurate and efficient license plate recognition systems, ultimately improving the efficiency and convenience of smart parking management in urban environments.

The main research contributions in this study are listed as follows:

- The study identifies the existing limitations in real-time license plate recognition for smart parking systems and highlights the need for more efficient and accurate methods to address these challenges.
- Generating an extensive custom dataset for car license plate recognition.
- The study proposes a YOLO-based approach with a custom dataset generation to improve license plate detection and recognition in smart parking systems, offering an effective solution for real-time processing and enhanced accuracy.
- The study conducts comprehensive experimental evaluations to validate the proposed method, assessing its performance in terms of accuracy, speed, and robustness. The results demonstrate the effectiveness and superiority of the YOLO-based approach in addressing the research gap and meeting the requirements of smart parking systems.

## II. PREVIOUS STUDIES REVIEW

The authors in [4] focuses on parking entrance control using license plate detection and recognition. The method employs computer vision techniques to detect and recognize license plates of vehicles entering a parking area. The advantages of this approach include enhanced security and improved efficiency in managing parking access. By automating the entrance control process, it reduces the need for manual intervention and enables real-time monitoring. However, potential drawbacks may include challenges in accurately detecting and recognizing license plates under varying lighting conditions and vehicle speeds. Overall, this research presents a practical solution for parking entrance control, offering benefits in terms of security and efficiency while acknowledging the need for robust license plate detection and recognition algorithms.

In [13], license plate recognition algorithms are explored based on deep learning in complex environments. The proposed method utilizes deep learning techniques to address challenges such as varying lighting conditions, occlusions, and plate deformations. The advantages of this approach include achieving high accuracy rates and robustness in complex scenarios, thanks to the superior feature extraction and pattern recognition capabilities of deep learning algorithms. However, drawbacks include the need for large amounts of labeled data for training, computational complexity, and potential limitations in generalizing to different license plate formats and languages. Overall, this research highlights the effectiveness of deep learning in license plate recognition but emphasizes the importance of considering data requirements, computational complexity, and adaptability to practical implementation.

The authors in [14] proposes a license plate recognition system that utilizes YOLOv5 and CNN (Convolutional Neural Network) models. The method involves training the YOLOv5 model to detect license plates in images, followed by using a CNN model for character recognition. The advantages of this approach include the ability to accurately locate and recognize license plates in real-time, enabling efficient automation of tasks such as parking management and access control. The use of deep learning models like YOLOv5 and CNN allows for robust performance even in challenging scenarios. However, potential drawbacks may include the need for a large amount of labeled data for training the models and the computational complexity associated with deep learning algorithms. Overall, this research demonstrates the effectiveness of using YOLOv5 and CNN for license plate recognition, offering advantages in terms of accuracy and real-time processing while acknowledging the considerations related to data requirements and computational resources.

The authors in [15] developed a low-cost IoT-based Arabic license plate recognition model for smart parking systems. The method utilizes existing IoT infrastructure and image processing algorithms to automate the recognition of Arabic license plates, enhancing parking management efficiency. The advantages include its cost-effectiveness and focus on Arabic license plates, catering to regions with Arabic as the dominant language. However, limitations include reliance on IoT infrastructure and the restriction to Arabic license plates. Challenges such as connectivity issues and varying plate formats may affect accuracy.

The authors in [17] presented an all-encompassing automated license plate recognition (ALPR) system, which streamlines the entire process by employing the YOLO (You Only Look Once) algorithm for vehicle and license plate detection, coupled with vehicle classification. The proposed approach represents a comprehensive solution to license plate recognition, aiming to eliminate the need for multiple separate algorithms and enhance the automation of this crucial task in various applications. However, the study brings to light a significant concern regarding the system's low accuracy rate in certain scenarios. This limitation suggests that the system may encounter challenges in reliably recognizing license plates, particularly in complex or adverse conditions. Consequently, it underscores the necessity for further research and improvements in accuracy to ensure the system's robust performance across diverse real-world scenarios and applications.

The study [18] explored a convolutional neural network (CNN)-based approach for license plate detection. The proposed method in the study centers around the utilization of CNNs, a class of deep learning algorithms, to automatically detect license plates in images. The authors conduct a comprehensive investigation to assess the effectiveness of this approach. Their findings indicate that the CNN-based method demonstrates promise in the realm of license plate detection, providing notable advantages in terms of speed and efficiency compared to traditional methods. However, a critical limitation that emerges from their research is a relatively low accuracy rate. This limitation is particularly evident in challenging scenarios, such as low-light conditions or when license plates

exhibit variations in size, angle, or perspective. The low accuracy rate underscores the need for further refinement and optimization of the CNN-based approach to improve its robustness and reliability, ensuring accurate license plate detection across diverse real-world situations and enhancing its potential applications in fields such as automated surveillance and transportation systems.

The study [19] presented a License Plate Recognition System (LPRS) utilizing an improved YOLOv5 (You Only Look Once) architecture in conjunction with a GRU (Gated Recurrent Unit) network. The proposed method aims to enhance the accuracy and efficiency of license plate recognition, a vital task with applications in security and traffic management. Through a comprehensive exploration of the proposed approach, the study reveals that the improved YOLOv5 and GRU-based LPRS system exhibits promising potential in recognizing license plates from images and videos. However, a significant limitation is identified, notably a lower accuracy rate in challenging conditions and scenarios, such as low lighting or extreme angles. This limitation underscores the need for further refinements and enhancements to address these challenges and improve the overall accuracy and robustness of the system, as accurate license plate recognition is crucial for various real-world applications, including surveillance and automated toll collection systems.

### III. MATERIAL AND METHOD

#### A. Dataset Generation

In this study, a custom dataset is generated for license plate recognition. For this dataset, following steps are performed.

1) *Collect images*: In this study, for collecting license plate images from internet resources to generate a custom dataset, several steps are performed. Firstly, we identify reliable sources that provide license plate images, including open data repositories and publicly available datasets. Secondly, systematically search and retrieve license plate images from these sources using specific keywords or filters with image databases that contain labeled license plate images. Thirdly, verify the authenticity and quality of the collected images to ensure they are suitable for dataset creation for image resolution, clarity, and visibility of the license plates. Finally, we consider verifying the accuracy and reliability of the labeled license plate information.

2) *Image annotation*: In collected images, there are several images with no labeling which are required to label and annotate. Annotating unlabeled license plate images for our custom dataset involves manually adding annotations to identify and localize license plate regions. The annotator uses specialized software to draw bounding boxes around the license plates and assign labels representing alphanumeric characters. Accuracy and attention to detail are crucial in ensuring precise annotations. Consistent annotation guidelines and regular quality checks maintain dataset quality. Annotated license plate images enable the training and evaluation of

accurate and robust license plate recognition models for real-world applications.

3) *Data augmentation*: To cover more variations in images and extending the dataset, we performed data augmentation. Augmenting license plate images is performed to expanding the custom dataset by applying various transformations and modifications to the existing images. This augmentation technique enhances the dataset's diversity and variability, making the trained model more robust and capable of handling different scenarios. Common augmentation techniques include rotation, translation, scaling, flipping, adding noise, changing lighting conditions, and introducing occlusions. By applying these transformations, the dataset can capture variations in license plate positions, angles, sizes, and image characteristics, such as brightness and noise levels. Augmentation helps to mitigate overfitting and improve the generalization ability of the license plate recognition model. It also enables the model to better handle real-world challenges, such as variations in weather, lighting conditions, and vehicle orientations. Therefore, we generated 8642 images with above steps in our custom dataset.

#### B. One-Stage Detector

One-stage detector algorithms are a type of object detection algorithm used in computer vision tasks. Unlike two-stage detectors that involve a separate region proposal network (RPN) and object detection stage, one-stage detectors perform both object localization and classification in a single pass. One popular example of a one-stage detector algorithm is the YOLO (You Only Look Once) algorithm. This algorithm enables to process the entire image in a single pass makes it efficient and suitable for real-time applications like license plate recognition in smart parking systems. Fig. 1 [16] shows the structure of one-stage detector.

In a one-stage detector algorithm, the input to the algorithm is an image that contains objects of interest, such as vehicles with license plates in the case of license plate recognition. The algorithm processes the entire image at once and produces a set of bounding boxes along with corresponding class predictions for each detected object. This is in contrast to two-stage detectors that first propose regions of interest and then classify those regions.

The backbone network is a crucial component of the one-stage detector algorithm. It is typically a deep convolutional neural network (CNN) that extracts hierarchical features from the input image. The backbone network can be a popular CNN architecture like ResNet, VGG, or Darknet, which has been pre-trained on a large-scale image classification task. The purpose of the backbone network is to capture high-level representations and semantic information from the image.

The neck network typically consists of additional convolutional layers that fuse and combine the multi-scale features from the backbone network. This fusion helps to enhance the representation power of the features and enables the algorithm to detect objects at different scales and resolutions.

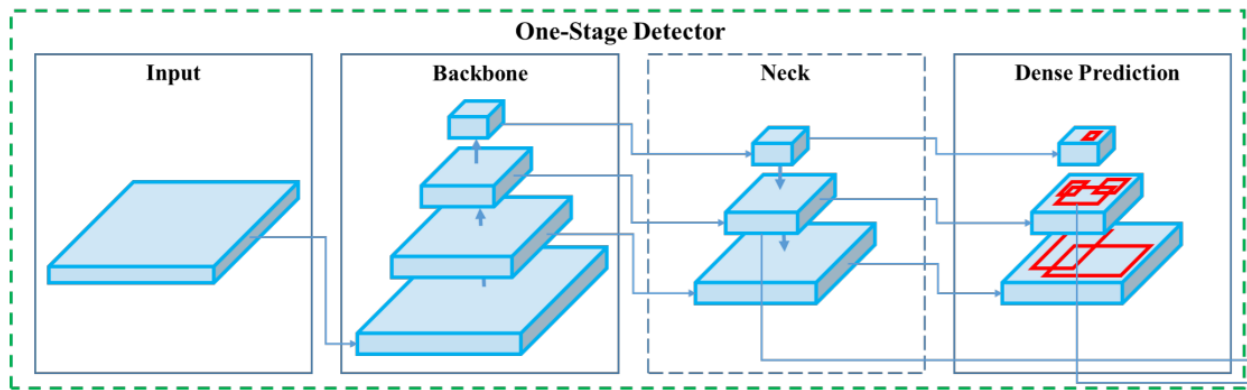


Fig. 1. Structure of one-stage detector.

The head or dense prediction part of the one-stage detector algorithm performs the final stage of object detection. It consists of convolutional layers followed by fully connected layers. The head network takes the fused features from the neck network as input and performs spatial and channel-wise convolutions to generate the final detection outputs. These outputs include the bounding box coordinates (x, y, width, height) and class probabilities for each detected object. The head network utilizes anchor boxes or default boxes to anchor the predicted bounding boxes to predefined scales and aspect ratios [16].

### C. Model Generation

Generating a YOLOv5 model for license plate recognition on a custom dataset involves several steps, including dataset preparation, training, validation, and testing. Firstly, the custom dataset is divided into training, validation, and testing sets with the typical split of 70%, 20%, and 10%, respectively. The training set, comprising 70% of the dataset, is used to train the YOLOv5 model. The validation set, consisting of 20% of the dataset, is utilized for monitoring the model's performance during training and tuning hyperparameters. Lastly, the testing set, which makes up 10% of the dataset, is used to evaluate the final model's performance.

1) *Training*: In the training module, the YOLOv5 model is trained on the training set using the labeled license plate images. This is typically achieved through an iterative process using techniques such as stochastic gradient descent (SGD) or

adaptive optimization algorithms. During training, the YOLOv5 model learns to detect and localize license plates in images and predict the corresponding alphanumeric characters. In training a YOLOv5 model for license plate recognition, several loss curves are commonly monitored during the training process: train/box\_loss, train/obj\_loss, and train/cls\_loss. Fig. 2 shows training loss curves.

As shown in Fig. 2, the train/box\_loss curve represents the loss associated with the bounding box predictions. The model aims to accurately predict the coordinates and dimensions of the bounding boxes around license plates. During training, the box loss gradually decreases as the model learns to better localize and fit the bounding boxes around the license plates. A lower box loss indicates improved precision and accuracy in the model's ability to detect and locate license plates.

The train/obj\_loss curve captures the loss related to objectness prediction. In YOLOv5, each grid cell predicts whether an object is present within it or not. The objectness loss calculates the difference between the predicted objectness scores and the ground truth labels. This loss helps the model to discriminate between objects and background regions. As the training progresses, the model learns to assign higher objectness scores to grid cells containing license plates and lower scores to empty cells, resulting in a decrease in the objectness loss. A decreasing obj\_loss curve indicates the model's improved capability to identify relevant regions for license plate recognition.



Fig. 2. Training loss curves.

The train/cls\_loss curve represents the loss associated with the classification of license plate characters. In license plate recognition, the model needs to classify the alphanumeric characters present on the license plates. The cls\_loss measures the dissimilarity between the predicted class probabilities and the ground truth character labels. The model aims to accurately recognize and classify the characters. As the training advances, the cls\_loss decreases, indicating that the model becomes more proficient at correctly identifying the characters on the license plates. A diminishing cls\_loss curve reflects the model's improved ability to perform accurate character classification.

By monitoring the train/box\_loss, train/obj\_loss, and train/cls\_loss curves during training, one can gain insights into the model's learning progress and performance in different aspects of license plate recognition. The decreasing trends in these loss curves indicate the model's improvement in bounding box prediction, objectness estimation, and character classification, respectively.

2) *Validation:* In the validation module, the YOLOv5 model is evaluated on the validation set to assess its

performance and fine-tune its hyperparameters. Based on the validation results, adjustments can be made to the model architecture, training parameters, or data augmentation techniques to improve its performance. During the validation process of a YOLOv5 model for license plate recognition, the loss curves serve as valuable metrics to assess the model's performance. Fig. 3 shows validation loss curves.

As shown in Fig. 3, the val/obj\_loss curve corresponds to the loss related to objectness prediction during validation. It evaluates the difference between the predicted objectness scores and the ground truth labels for license plates in the validation images. As the model learns to distinguish between objects and background regions, the val/obj\_loss decreases. A decreasing obj\_loss curve indicates that the model is becoming more proficient at identifying relevant regions for license plate recognition during validation. This demonstrates the model's improved capability to assign higher objectness scores to grid cells containing license plates and lower scores to empty cells, enhancing its ability to discriminate between objects and background regions.

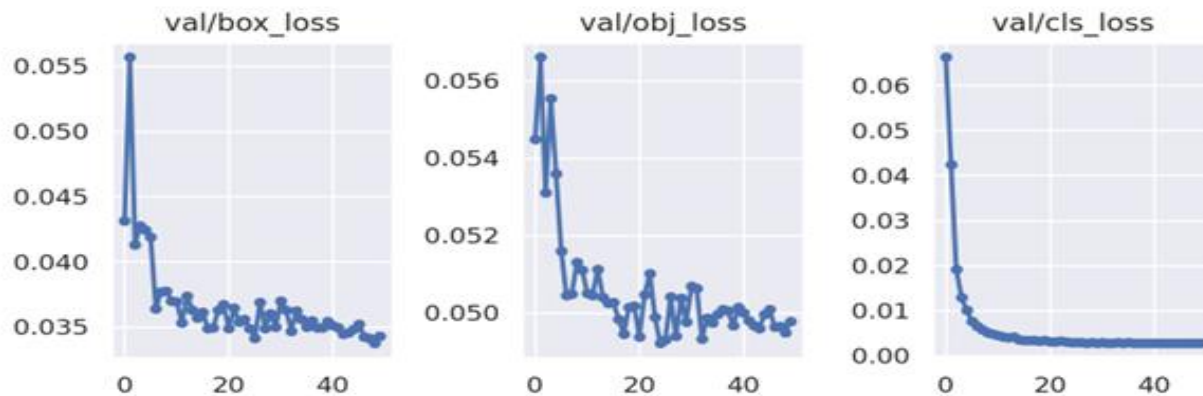


Fig. 3. Validation loss curves.

The val/cls\_loss curve represents the loss associated with the classification of license plate characters during validation. The model aims to accurately recognize and classify the alphanumeric characters. As the model improves its ability to correctly identify characters on license plates during validation, the val/cls\_loss, decreases. A diminishing cls\_loss curve indicates that the model becomes more proficient at accurately classifying characters on license plates. This signifies the model's enhanced ability to perform accurate character recognition and classification during validation.

By monitoring the val/box\_loss, val/obj\_loss, and val/cls\_loss curves during the validation process, one can evaluate the model's performance in bounding box prediction, objectness estimation, and character classification for license plate recognition tasks. Decreasing trends in these loss curves indicate the model's improvement in accurately localizing license plates, discriminating between objects and background regions, and correctly identifying characters.

#### IV. EXPERIMENTAL RESULTS

This section presents experimental results and performance evaluation. For these results, the trained YOLOv5 model is

applied to the unseen license plate images in the testing set. The model's predictions are compared against the ground truth annotations to evaluate its performance on new and unseen data. These metrics are precision, recall and F1-score are computed to assess the model's effectiveness in license plate recognition. We use the testing sets for this evaluation. The testing module provides insights into the model's generalization ability and its performance in real-world scenarios. By following this three-module approach, the YOLOv5 model is trained on the labeled license plate images from the training set, validated on the validation set to fine-tune its performance, and then tested on the unseen license plate images from the testing set to evaluate its effectiveness in license plate recognition tasks. This process ensures the development of a robust and accurate YOLOv5 model for license plate recognition on the custom dataset.

##### A. Confusion Matrix

Confusion matrix as popular tool for evaluating the performance of deep learning-based models is utilized in this experiment. It serves as a visual representation and quantitative assessment of the license plate recognition YOLOv5 model's



predictive performance, allowing to evaluate, analyze, and improve the model's accuracy and effectiveness. Additionally, the confusion matrix enables the calculation of class-specific performance metrics, providing a more detailed understanding of the model's strengths and weaknesses across different license plate character classes.

As shown in Fig. 4, the confusion matrix is organized as a grid, with the Y-axis representing the predicted classifications and the X-axis representing the true classifications. The main diagonal of the confusion matrix corresponds to the correctly classified instances, where the predicted class matches the true class. Off-diagonal cells represent the misclassified instances, where the predicted class differs from the true class.

### B. Performance Measurement

By analyzing the confusion matrix, performance measurements can be performed. Popular metrics such as precision, recall, F1-score and mAP are calculated from the values in the confusion matrix.

1) *Precision-confidence curve*: The precision-confidence curve is a visual representation of the relationship between precision and confidence threshold in a license plate recognition YOLOv5 model. The precision-confidence curve plots the precision value at different confidence thresholds. Starting from a high confidence threshold, the precision will be relatively high because only confident predictions are considered valid. Fig. 5 demonstrates the precision-confidence curve.

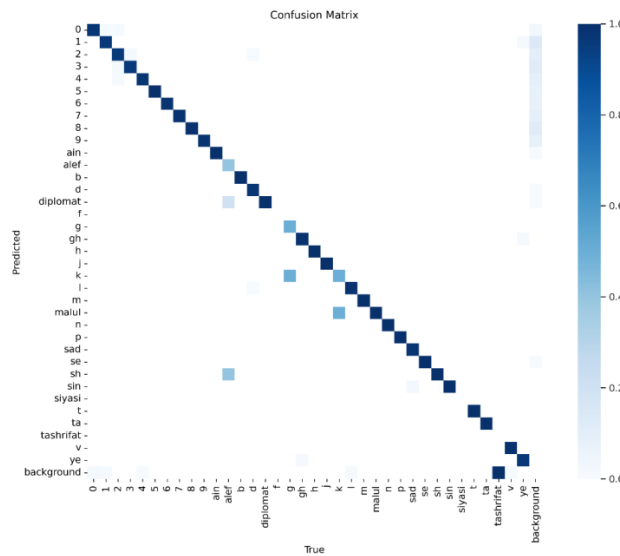


Fig. 4. Results of confusion matrix.

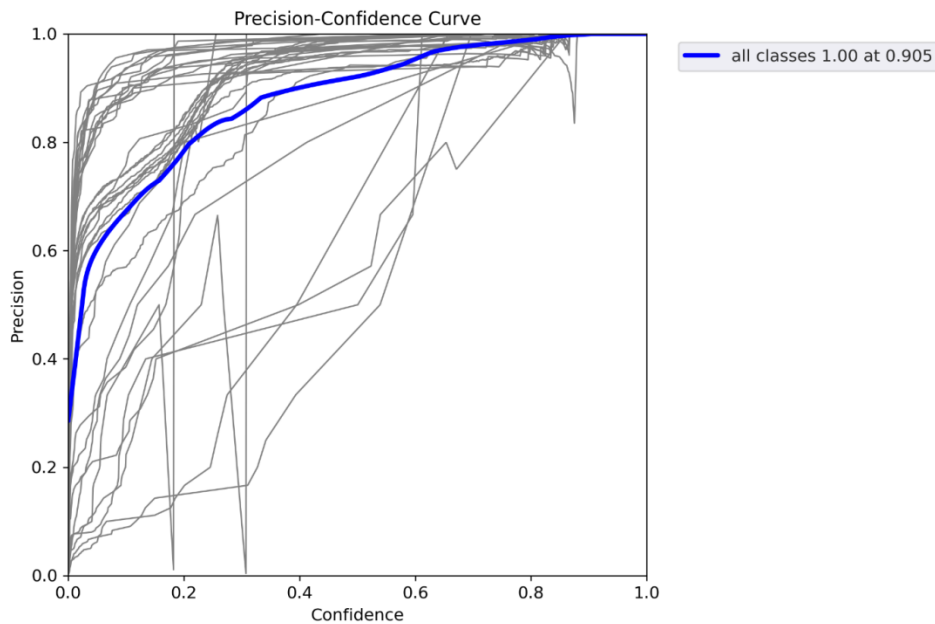


Fig. 5. The precision-confidence curve.

As shown in Fig. 5, The X-axis represents the confidence threshold, while the Y-axis represents the precision. The curve demonstrates how precision changes as the confidence threshold is adjusted. Initially, with a high confidence threshold, the precision is high because only confident predictions are considered valid. As the confidence threshold decreases, more predictions are included, potentially leading to false positives and a decrease in precision. Higher precision implies a lower false positive rate, indicating fewer incorrect predictions. However, setting a high confidence threshold may result in missing true positive instances, reducing efficiency. By analyzing the curve, it is possible to identify the optimal confidence threshold that strikes a balance between precision

and efficiency. This information enables to assess the model's performance and adjust the confidence threshold to achieve the desired accuracy and recognition efficiency in the license plate recognition model.

2) *Recall-confidence curve*: The recall-confidence curve illustrates how the recall metric changes as the confidence threshold varies. As the confidence threshold decreases, more predictions are considered valid, including both true positives and potential false positives. Fig. 6 shows the recall-confidence curve.

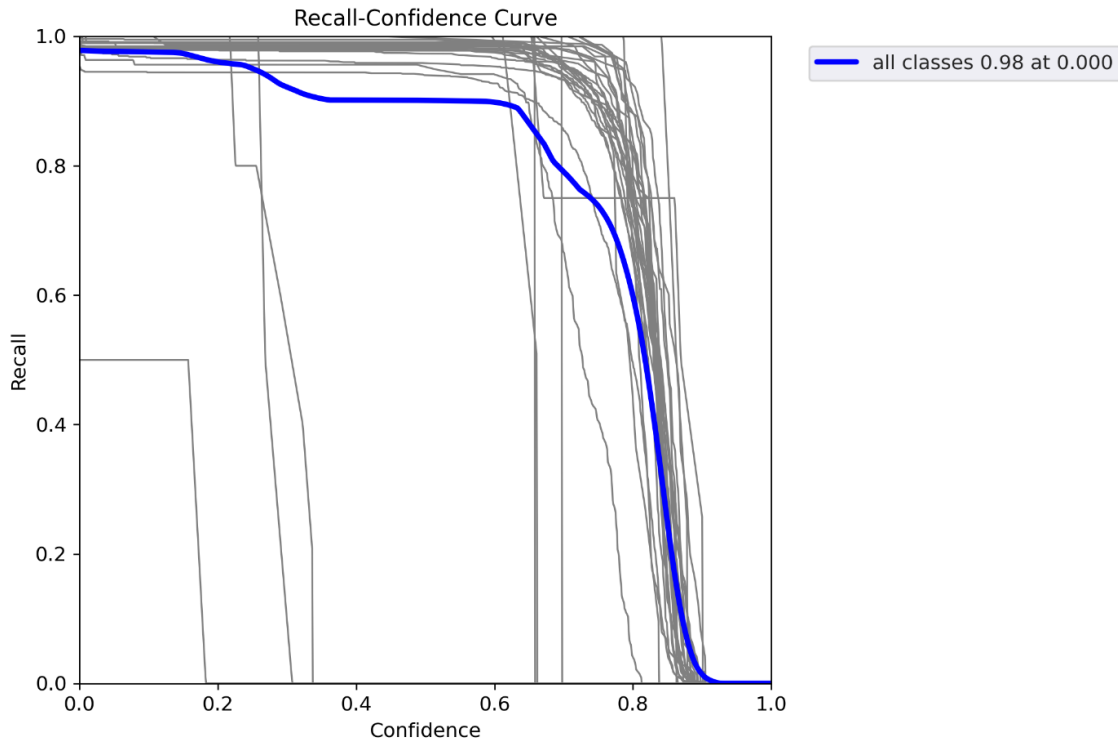


Fig. 6. The recall-confidence curve.

As shown in Fig. 6, the X-axis of the curve represents the confidence threshold, which determines the minimum confidence score required for a prediction to be considered valid. The Y-axis represents the recall, also known as sensitivity or true positive rate. Recall measures the proportion of correctly predicted positive instances (true positives) out of all actual positive instances (true positives plus false negatives). Higher recall indicates that the model is better at capturing and recognizing actual positive instances, minimizing false negatives. However, a lower confidence threshold that maximizes recall may also introduce more false positives, reducing the model's efficiency. It becomes a trade-off between correctly identifying more positive instances and the potential inclusion of false positives.

3) *Precision-Recall curve*: The precision-recall curve is another visual representation that provides insights into the performance and efficiency of a license plate recognition YOLOv5 model. The curve showcases the relationship

between precision and recall, two important metrics used to evaluate the model's predictive accuracy. Fig. 7 shows the precision-recall curve.

The X-axis of the precision-recall curve represents the recall, also known as sensitivity or true positive rate. The Y-axis represents precision, which indicates the proportion of correctly predicted positive instances out of all instances predicted as positive (true positives plus false positives). The precision-recall curve illustrates how the precision metric change as the recall varies. A higher recall indicates that the model is capturing a larger number of positive instances, minimizing false negatives. As the recall increases, the model is becoming more sensitive and successfully identifying more actual positive instances. However, as the recall increases, it becomes more challenging to maintain a high precision. The inclusion of more positive instances may introduce false positives, impacting the precision of the model.

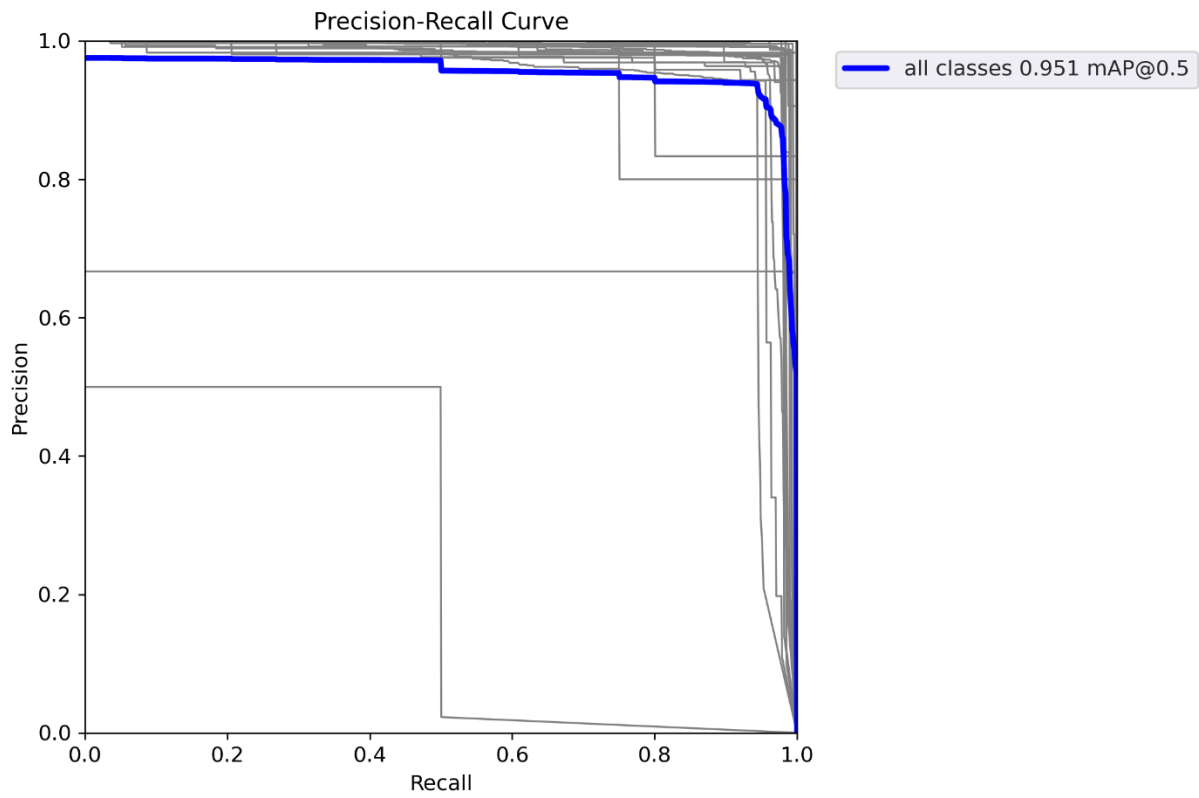


Fig. 7. The precision-recall curve.

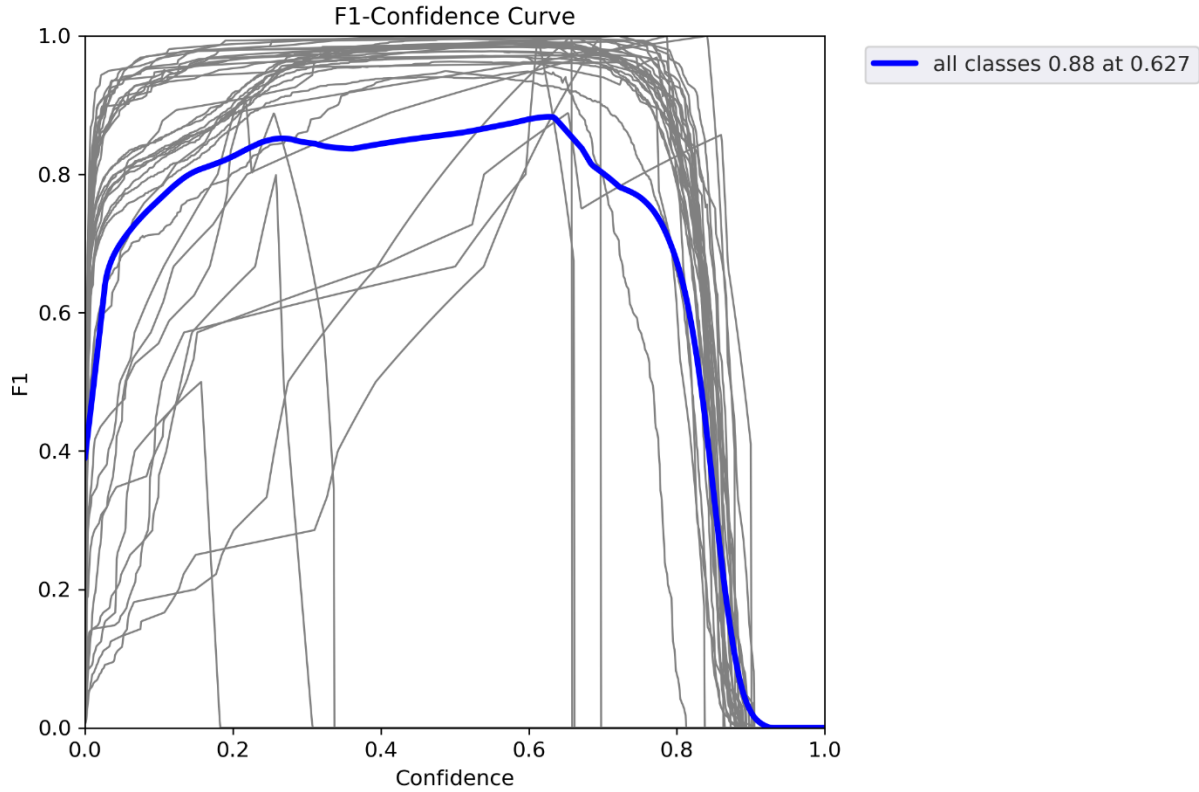


Fig. 8. F1-score curve.

4) *F1-score*: The F1-score curve demonstrates how the F1-score change as the threshold varies. The F1-score is important for measuring the efficiency of a license plate recognition YOLOv5 model because it combines precision and recall into a single value. It provides a comprehensive assessment of the model's ability to accurately identify positive instances while avoiding false positives.

As shown in Fig. 8, the curve illustrates the relationship between the F1-score and a varying threshold used for predictions. The X-axis of the F1-score curve represents the threshold, which is the minimum confidence score required for a prediction to be considered valid. The Y-axis represents the F1-score, which is a metric that combines both precision and recall into a single value. The F1-score provides a balanced evaluation of the model's performance, taking into account both the ability to correctly identify positive instances (precision) and the ability to capture all actual positive instances (recall). As the threshold decreases, more predictions are considered valid, potentially increasing the recall of the model.

## V. CONCLUSION

In this study, we introduce a YOLO-based method complemented by a custom dataset tailored for license plate recognition within smart parking systems. Our research effectively addresses the primary research gap, delivering real-time processing capabilities and enhanced accuracy. The extensive custom dataset empowers the model to handle even the most challenging scenarios frequently encountered in smart parking environments. Our comprehensive experimental evaluations underscore the method's superior performance across accuracy, processing speed, and robustness. The significant contributions of this research encompass the identification of limitations, the creation of a specialized dataset, the proposal of the YOLO-based approach, and the thoroughness of our evaluations. These findings collectively advance the development of precise and efficient license plate recognition systems, thereby augmenting the management of smart parking in urban areas. It's worth noting that our custom dataset specifically targets the license plate types prevalent in the studied region or environment. Nevertheless, variations in license plate designs, formats, fonts, sizes, colors, and layouts exist across different countries, regions, or vehicle types. Future endeavors can prioritize the development of methods capable of generalizing and adapting to this diversity, potentially incorporating transfer learning, data augmentation techniques, or domain adaptation strategies to enhance the model's proficiency in accurately recognizing and interpreting various license plate variations.

## REFERENCES

- [1] T. Perković, P. Šolić, H. Zargariasl, D. Čoko, and J. J. Rodrigues, "Smart parking sensors: State of the art and performance evaluation," *Journal of Cleaner Production*, vol. 262, p. 121181, 2020.
- [2] C. Biyik et al., "Smart parking systems: Reviewing the literature, architecture and ways forward," *Smart Cities*, vol. 4, no. 2, pp. 623-642, 2021.
- [3] R. Nawaratne, S. Kahawala, S. Nguyen, and D. De Silva, "A generative latent space approach for real-time road surveillance in smart cities," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 7, pp. 4872-4881, 2020.
- [4] M. S. Farag, M. M. El Din, and H. El Shenbary, "Parking entrance control using license plate detection and recognition," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 15, no. 1, pp. 476-483, 2019.
- [5] S. M. Mutua, "An automatic number plate recognition system for car park management," *Strathmore University*, 2016.
- [6] A. Fahim, M. Hasan, and M. A. Chowdhury, "Smart parking systems: comprehensive review based on various aspects," *Heliyon*, vol. 7, no. 5, p. e07050, 2021.
- [7] G. T. Sutar, A. M. Lohar, and P. M. Jadhav, *Number plate recognition using an improved segmentation*. Citeseer, 2019.
- [8] S. M. Silva and C. R. Jung, "Real-time license plate detection and recognition using deep convolutional neural networks," *Journal of Visual Communication and Image Representation*, vol. 71, p. 102773, 2020.
- [9] A. Farley and H. Ham, "Real time IP camera parking occupancy detection using deep learning," *Procedia Computer Science*, vol. 179, pp. 606-614, 2021.
- [10] R. N. Babu, V. Sowmya, and K. Soman, "Indian car number plate recognition using deep learning," in *2019 2nd international conference on intelligent computing, instrumentation and control technologies (ICICT)*, 2019, vol. 1: IEEE, pp. 1269-1272.
- [11] J. Du, "Understanding of object detection based on CNN family and YOLO," in *Journal of Physics: Conference Series*, 2018, vol. 1004: IOP Publishing, p. 012029.
- [12] S. Wu, X. Li, and X. Wang, "IoU-aware single-stage object detector for accurate localization," *Image and Vision Computing*, vol. 97, p. 103911, 2020.
- [13] W. Weihong and T. Jiaoyang, "Research on license plate recognition algorithms based on deep learning in complex environment," *IEEE Access*, vol. 8, pp. 91661-91675, 2020.
- [14] S. Raj, Y. Gupta, and R. Malhotra, "License plate recognition system using yolov5 and cnn," in *2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 2022, vol. 1: IEEE, pp. 372-377.
- [15] M. M. Abdellatif, N. H. Elshabasy, A. E. Elashmawy, and M. AbdelRaheem, "A low cost IoT-based Arabic license plate recognition model for smart parking systems," *Ain Shams Engineering Journal*, vol. 14, no. 6, p. 102178, 2023.
- [16] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "Yolov4: Optimal speed and accuracy of object detection," *arXiv preprint arXiv:2004.10934*, 2020.
- [17] Al-Batat, Reda, et al. "An end-to-end automated license plate recognition system using YOLO based vehicle and license plate detection with vehicle classification." *Sensors* 22.23 (2022): 9477.
- [18] Cao, Yong. "Investigation of a convolutional neural network-based approach for license plate detection." *Journal of Optics* (2023): 1-7.
- [19] Shi, Hengliang, and Dongnan Zhao. "License Plate Recognition System Based on Improved YOLOv5 and GRU." *IEEE Access* 11 (2023): 10429-10439.

# Enhancing Decision-Making with Data Science in the Internet of Things Environments

Lei Hu<sup>1\*</sup>, Yangxia Shu<sup>2</sup>

Operation and Maintenance Section of Assets Department, Jiangxi Institute of Fashion Technology  
Nanchang 330201, Jiangxi, China<sup>1</sup>

College of Big Data Science, Jiangxi Institute of Fashion Technology, Nanchang 330201, Jiangxi, China<sup>2</sup>  
Information Technology Integration Innovation Center, Intelligent Research and Innovation Team for Clothing  
(Jiangxi Institute of Fashion Technology), Nanchang 330201, Jiangxi, China<sup>1,2</sup>

Information Technology Integration Innovation Center, Intelligent Research and Innovation Team for Clothing  
(Jiangxi Institute of Fashion Technology), Nanchang 330201, Jiangxi, China<sup>2</sup>

**Abstract**—The Internet of Things (IoT) has emerged as a transformative technology, enabling various devices to interconnect and generate vast amounts of data. The insights contained within this data can revolutionize industries and improve decision-making processes. The heterogeneity, scale, and complexity of IoT data pose challenges for efficient analysis and utilization. In this paper, the field of data science is explored in the IoT context, focusing on critical techniques, applications, and challenges vital to realizing the full potential of IoT data. This paper explores the field of data science in the IoT context, focusing on critical techniques, applications, and challenges vital to realizing the full potential of IoT data. The distinctive qualities of IoT data, including its volume, velocity, variety, and veracity, are examined, and their impact on data science approaches is analyzed. Additionally, cutting-edge data science approaches and methodologies designed for IoT data, such as data preprocessing, data fusion, machine learning, and anomaly detection, are discussed. The importance of scalable and distributed data processing frameworks to handle IoT data's large-scale and real-time nature is highlighted. Furthermore, the application of data science in various IoT fields, such as smart cities, healthcare, agriculture, and industrial IoT, is explored. Finally, areas for future research and development are identified, such as privacy and security issues, understanding machine learning models, and ethical aspects of data science in IoT.

**Keywords**—Internet of Things; IoT data; data science; data preprocessing; machine learning; real-time analytics

## I. INTRODUCTION

### A. Background and Motivation

The Internet of Things (IoT) creates a world where the objects around people can sense and gather information about the environment [1]. With the proliferation of IoT devices in diverse domains such as smart homes, healthcare, transportation, and industrial systems, an enormous amount of data is continuously generated. This data presents immense potential for extracting valuable insights and driving informed decision-making [2]. However, harnessing the full potential of IoT data requires effective data science techniques and approaches [3]. This is where the significance of meta-heuristic algorithms, Machine Learning (ML), deep learning, Artificial Intelligence (AI), and urban public transportation becomes apparent in the context of data science in IoT environments.

Meta-heuristic algorithms are vital tools within the data science toolkit, as they provide intelligent, heuristic-based optimization techniques for solving complex problems. In the realm of IoT, these algorithms can be employed to optimize resource allocation, enhance data processing efficiency, and address challenges related to data routing, sensor placement, and energy management [4]. ML, a subset of AI, is at the forefront of IoT data analysis. ML algorithms empower IoT applications to learn from historical data, recognize patterns, and make predictions or decisions autonomously. They are instrumental in understanding the behavior of connected devices, detecting anomalies, and predicting future trends within IoT ecosystems [5, 6]. Deep learning, a subfield of ML, has gained substantial importance in IoT data analysis due to its ability to handle large-scale, unstructured data [7]. Deep neural networks excel in feature extraction and abstraction, making them invaluable for image and speech recognition in IoT applications such as surveillance and voice-controlled devices [8-10]. AI, encompassing ML and deep learning, extends the capabilities of IoT by enabling devices to exhibit human-like intelligence. This manifests in autonomous decision-making, natural language processing, and adaptive behavior, empowering IoT systems to become more responsive, efficient, and user-friendly [11]. Urban public transportation systems are a crucial domain within the IoT landscape. IoT sensors and data science techniques are instrumental in optimizing public transportation networks, reducing congestion, improving routing efficiency, and enhancing the overall commuter experience. Real-time data analytics, enabled by IoT and data science, can transform urban mobility and contribute to sustainability efforts [12].

The motivation behind this paper is twofold. Firstly, the rapid growth of IoT devices and the resulting data deluge present unique challenges in data management, processing, and analysis. The sheer volume, velocity, and variety of IoT data require advanced data science techniques capable of handling and extracting meaningful information from this data. Therefore, there is a need to explore and develop specialized data science techniques tailored to the unique characteristics of IoT data. Secondly, the application of data science in the IoT domain holds significant potential for driving innovation and creating value. Organizations can uncover hidden patterns, detect anomalies, optimize operations, and enhance decision-

making in IoT-based systems by leveraging data science techniques. This has implications for various sectors, including healthcare, energy management, environmental monitoring, and smart cities, where data-driven insights can improve efficiency, sustainability, and quality of life.

### B. Objectives and Scope

This paper aims to examine the application of data science techniques in the context of IoT and assess the potential benefits and challenges involved. The paper aims to achieve the following specific objectives:

- Examining the current state of data science techniques and methodologies and their relevance to IoT data analysis.
- Identifying the challenges and limitations of applying data science in the IoT domain, such as data heterogeneity, scalability, real-time processing, and privacy concerns.
- Exploring using ML algorithms, statistical analysis, and data mining techniques for extracting meaningful insights from IoT data.
- Investigating the integration of IoT data with other data sources, such as social media, weather data, and sensor networks.
- Evaluating the performance and effectiveness of data science techniques in real-world IoT applications through case studies and experiments.
- Discussing the ethical and privacy implications associated with collecting, storing, and analyzing IoT data.

The scope of this review paper encompasses a comprehensive analysis of data science techniques and their application in the IoT domain. It covers various topics, including data preprocessing and cleaning, feature engineering, anomaly detection, predictive modeling, and visualization techniques tailored for IoT data. The paper considers supervised and unsupervised learning algorithms and advanced techniques like deep learning and ensemble methods. Additionally, it explores the challenges of handling high-dimensional, streaming, and heterogeneous IoT data and proposes solutions to address these challenges. The paper focuses on the practical implications of using data science for the IoT. It examines real-world case studies and applications to highlight data science techniques' potential benefits and limitations in diverse IoT domains such as smart cities, healthcare monitoring, industrial automation, and environmental sensing. Furthermore, the paper acknowledges the ethical considerations and privacy concerns associated with IoT data collection and analysis, providing insights into responsible data practices and regulatory frameworks.

### C. Organization of the paper

The paper is organized as follows: Section II discusses the challenges and considerations of analyzing IoT data using data science methods. Section III delves into the various data science techniques that can be utilized for IoT data analysis.

Scalable data processing for IoT data is discussed in Section IV. This section also explores real-world data science applications in the IoT domain, showcasing successful case studies and their outcomes. In Section V, we present open research challenges and future directions. Finally, in Section VI, the conclusion of the paper is provided.

## II. BACKGROUNDS

This section delves into the unique challenges posed by IoT data that impact the application of data science techniques. It explores the specific characteristics of IoT data, including its sheer volume, velocity, and variety. The section discusses the inherent complexity of IoT data, such as its unstructured nature, real-time streaming nature, and potential for high dimensionality. Furthermore, it highlights the issues related to data quality, including missing values, noise, and inconsistencies, which can pose significant challenges for data science practitioners. The section also addresses the security and privacy concerns associated with IoT data, emphasizing the need for robust data protection mechanisms. Additionally, it explores the issue of data interoperability, as IoT devices often use different data formats and protocols, making data integration and analysis more challenging.

### A. IoT Data Characteristics

As shown in Fig. 1, the field of data science faces specific challenges when dealing with IoT data, mainly related to volume, velocity, variety, veracity, value, and variability. The sheer volume of data generated by IoT devices is immense. With billions of interconnected devices, the amount of data produced exponentially increases. This massive volume of data poses storage, processing, and analysis challenges [13]. IoT data is generated in real-time or near real-time, often streaming continuously from various sources. This high velocity of data requires data scientists to implement real-time analytics solutions that can process and analyze data on the fly [14]. IoT data comes in diverse formats and types. It includes structured data from sensors, unstructured data such as images and videos, and text data from social media. The variety of IoT data poses integration, quality, and interoperability challenges [15]. The veracity of IoT data refers to its reliability, accuracy, and trustworthiness. IoT data is often prone to errors, noise, and inconsistencies due to device malfunctions, network issues, or data transmission errors [16]. The data generated by IoT objects holds immense value in optimizing applications and uncovering novel insights and knowledge [17]. The speed of data collection from IoT devices can vary depending on the events triggering data collection, such as shopping data. Furthermore, data format changes may occur when devices are replaced or updated [18].

### B. IoT data properties

IoT data can be categorized based on spatial, temporal, and sensing properties. Each property plays a significant role in understanding and analyzing IoT data. A summary of these properties is presented in Table I. In the analysis of IoT data, the spatial properties of the data are diverse and have significant implications for selecting appropriate systems and techniques. One important consideration is whether the IoT devices are fixed or mobile [19]. For instance, environmental sensors in street lamps have fixed spatial information, while

vehicles possess mobile spatial information. This distinction is crucial as it impacts the analysis methods and the interpretation of the collected data. Another aspect to consider is the shape of the IoT data. Spatial data can be points, areas, line strings, or multiple disconnected points and areas. For example, weather information may be represented by polygons, whereas road networks are typically depicted using line strings.

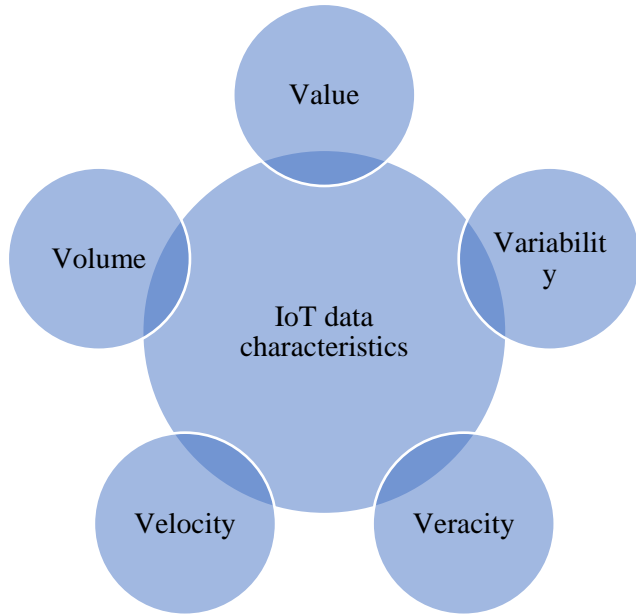


Fig. 1. IoT data characteristics.

Furthermore, it is crucial to consider any constraints imposed on the spatial information. For example, specific IoT applications involve restrictions on movement, such as cars being restricted to roads. Dealing with these constraints often requires preprocessing techniques to ensure accurate analytical results and avoid misinterpretations. Temporal information is another critical aspect of IoT data that needs to be considered. One key consideration is the nature of data updates. It is essential to determine whether data updates occur periodically or sporadically [20]. If data updates are periodic, it indicates that there are efficient methods for storing and load-balancing IoT data for stream processing. The data can be easily partitioned without any temporal skew in such cases.

On the other hand, if data updates are irregular and do not follow a specific pattern, additional techniques are required to partition the IoT data effectively and avoid temporal skew. Temporal skew can lead to imbalances in data processing and analysis, impacting the accuracy and timeliness of results.

Another temporal property to consider is the update frequency of the data. This frequency can vary significantly, ranging from frequent updates to less frequent ones. Data compaction techniques can be employed effectively when data updates occur periodically and frequently. Data compaction involves reducing the volume of data by storing only the same or similar data, as they are likely to be obtained in subsequent updates. This approach can help optimize storage and processing efficiency.

Sensing values play a significant role in IoT data, providing crucial information about the physical world. These values exhibit various types, including numerical values, text, labels, and images [21]. However, due to the independent deployment of IoT devices by different organizations, the names and units assigned to attributes can vary, even when sensed by the same type of data. For instance, one device may measure temperature and label it as "temperature," while another may use "temp" as the attribute name. Similarly, units of measurement can differ, such as Celsius or Fahrenheit. In addition to variations in attribute names and units, sensing quality is another essential characteristic of sensing values. Some sensors are highly accurate, providing reliable and precise measurements. On the other hand, specific sensors may be less accurate, leading to potential noise, errors, or invalid data in the collected values. IoT systems must have mechanisms in place to handle such incorrect values appropriately. These mechanisms may involve data cleansing, filtering, or error detection techniques to ensure the integrity and reliability of the analyzed IoT data.

### C. IoT Data Examples

In the smart city data analysis field, various data sources are utilized. These examples highlight some of the common types of IoT data analyzed in the context of smart cities [22]. Drone (UAV) data is increasingly used in smart city applications for various purposes, including environmental monitoring and surveillance [23]. The spatial representation of drone data is typically in the form of 3-D trajectories, capturing the movement of drones in the airspace [24]. Temporally, updates are periodic but occur at a low frequency. Sensing values in this context include air quality indicators such as PM2.5 and CO2 levels [25]. Shopping data provides insights into consumer behavior and trends within a smart city. Spatially, shopping data is often represented as fixed 2-D points, such as locations of retail stores or shopping centers. Temporally, updates are non-periodic and occur at a low frequency. Sensing values associated with shopping data include sales and user data, which can be used for market analysis and personalized marketing strategies [26].

TABLE I. IoT DATA PROPERTIES

Sensing	Temporal	Spatial
Quantity Unit Schema Data type	Update frequency Periodic or non-periodic	Constraint Relative or absolute Dimension Spatial shape Mobile or fixed

Public vehicle data is another key source of information in smart city analysis. This type of data captures the movements and behaviors of vehicles within a city [27]. The spatial aspect of public vehicle data is typically represented as 2-D trajectories along roads. Temporally updates are often periodic and occur at a high frequency. Sensing values in this context can include WiFi signals, environmental data, and vehicle-specific information such as speed and wheel speed. Weather data is an essential aspect of smart city analysis, providing insights into spatial patterns of temperature, humidity, and weather conditions. The spatial representation of weather data is typically in the form of fixed 2-D areas, while periodic updates at a low frequency characterize the temporal aspect. Sensing values associated with weather data include temperature, humidity, and weather labels. These examples illustrate the diverse nature of IoT data that can be collected and analyzed in the context of smart cities. The data can be obtained from various IoT devices or retrieved from the web, contributing valuable insights for urban planning, resource optimization, and decision-making processes.

#### D. Data Acquisition and Preprocessing

The challenges of IoT data for data science extend beyond volume, velocity, variety, and veracity to include specific data acquisition and preprocessing issues. These challenges are critical as they directly impact the quality and reliability of the data used for analysis and decision-making.

1) *Data acquisition*: IoT devices generate massive amounts of data, but acquiring that data can be challenging. IoT devices are distributed across various locations and environments, making data collection complex and heterogeneous [28]. Data scientists must consider data access, compatibility, and synchronization factors when acquiring IoT data. They must establish reliable data acquisition mechanisms, such as data streams or APIs, to capture and collect data in real-time or at regular intervals.

2) *Data preprocessing*: IoT data often requires extensive preprocessing before analysis. The raw data obtained from IoT devices may contain missing values, outliers, noise, and inconsistencies [29]. Preprocessing techniques are essential to clean, transform, and prepare the data for analysis. Data scientists need to address data quality issues, handle missing values through imputation methods, detect and handle outliers, and perform data normalization or scaling to ensure the data is in a suitable format for analysis.

3) *Data fusion*: IoT data is typically generated from multiple sources, such as sensors, wearables, social media, etc [30]. Integrating and fusing data from diverse sources is a significant challenge. Data fusion techniques need to be applied to combine and integrate data from different sensors or devices, ensuring that the resulting dataset provides a comprehensive and accurate representation of the phenomenon under study. Data scientists must consider the data's semantic, temporal, and spatial aspects to fuse and integrate the information effectively.

4) *Data privacy and security*: IoT data often contain sensitive and personal information, raising concerns about

privacy and security [31]. Data scientists must adhere to privacy regulations and implement robust security measures to protect the confidentiality, integrity, and availability of IoT data. Anonymization techniques, encryption methods, and access control mechanisms are crucial to ensuring data privacy and preventing unauthorized access or data breaches.

By effectively addressing the challenges of data acquisition and preprocessing, data scientists can enhance the reliability and usability of IoT data. This, in turn, enables more accurate and insightful analysis, leading to informed decision-making and the development of innovative applications in various domains, including smart cities, healthcare, transportation, and more. Continued research and advancements in data acquisition and preprocessing techniques are vital to overcoming these challenges and leveraging the full potential of IoT data for data science applications.

#### E. Scalability and Real-Time Processing

Scalability and real-time processing are two critical challenges that arise due to the massive influx of data from IoT devices. IoT data is generated at an unprecedented scale [32]. As the number of connected devices continues to grow, the volume of data generated increases exponentially. Handling and analyzing such massive amounts of data poses scalability challenges for data scientists [33]. Traditional data processing approaches may not be sufficient to handle the scale of IoT data. Data scientists need to design and implement scalable architectures and algorithms that can efficiently process and analyze large-scale IoT datasets. This involves distributed computing techniques, parallel processing, and using cloud-based infrastructures to handle the computational and storage demands of IoT data analysis. IoT data is time-sensitive, and real-time processing is essential to extract timely insights and enable immediate actions. Many IoT applications, such as smart cities, industrial monitoring, and healthcare, require real-time analytics to detect anomalies, make predictions, and trigger automated responses. Real-time processing of IoT data involves handling high-velocity data streams and making rapid decisions based on the analyzed data. Data scientists must develop streaming data processing frameworks and real-time analytics models to address the continuous flow of IoT data and generate insights in near real-time. This requires efficient algorithms, event-processing techniques, and low-latency systems to process and analyze data as it arrives.

Addressing the scalability and real-time processing challenges of IoT data requires advanced technologies and techniques. Data scientists need to leverage distributed computing frameworks such as Apache Hadoop or Apache Spark for parallel processing and handling large-scale IoT datasets [34]. They must also adopt real-time streaming platforms like Apache Kafka or Apache Flink to take high-velocity data streams and perform real-time analytics. Additionally, ML and AI algorithms can be applied to develop predictive models and anomaly detection systems that operate in real-time. By addressing the challenges of scalability and real-time processing, data scientists can unlock the full potential of IoT data for timely and informed decision-making. Processing and analyzing IoT data at scale and in real-time enables proactive monitoring, predictive maintenance, and



rapid response to emerging events and trends. However, ongoing research and innovation are required to develop more efficient and scalable data processing frameworks, algorithms, and architectures to keep pace with the ever-growing influx of IoT data.

#### F. Privacy and Security Considerations

The challenges of IoT data for data science encompass the technical aspects and the critical concerns of privacy and security. The vast amount of data generated by IoT devices poses significant challenges in ensuring the privacy and security of sensitive information.

1) *Privacy*: IoT devices collect a wide range of personal and sensitive data, including location information, health data, and behavioral patterns. Preserving the privacy of individuals becomes a crucial challenge as this data is transmitted, stored, and processed [35]. Data scientists must implement privacy-preserving techniques such as data anonymization, encryption, and access controls to safeguard personal information. Additionally, they must comply with privacy regulations and frameworks, such as the General Data Protection Regulation (GDPR), to ensure IoT data's lawful and ethical handling.

2) *Security*: IoT devices are vulnerable to security threats due to their heterogeneous nature, limited resources, and broad deployment [36]. They can be susceptible to attacks such as unauthorized access, data breaches, and tampering. Data scientists must address the security challenges by implementing robust security mechanisms. This includes ensuring secure communication protocols, device authentication, data encryption, and intrusion detection systems. Continuous monitoring and threat intelligence are essential to identify and mitigate potential security risks.

3) *Data governance*: The diverse nature of IoT data, collected from various sources and devices, poses challenges regarding data quality, integrity, and reliability [37]. Data scientists need to establish effective data governance frameworks to address these challenges. This involves data validation, data cleansing, and verifying data quality standards to ensure the accuracy and reliability of IoT data. Additionally, data scientists must establish data access controls and implement data lifecycle management practices to manage data throughout its lifecycle, including data retention and secure data disposal.

4) *Ethical considerations*: Data scientists need to be aware of the ethical implications of collecting and analyzing massive amounts of IoT data [38]. They must ethically handle data, ensure the informed consent of individuals, avoid bias in data analysis, and maintain transparency in data processing practices. Adhering to ethical guidelines and frameworks helps build trust among users and promotes responsible and accountable use of IoT data.

Addressing privacy and security challenges requires a comprehensive approach involving technical measures, regulatory compliance, and ethical considerations. Data scientists should collaborate with experts in privacy and security to design and implement robust security architectures,

privacy-enhancing techniques, and privacy impact assessments. Additionally, raising awareness among users about the privacy implications of IoT data and providing transparent data handling practices can help build trust and confidence in using IoT technologies.

### III. DATA SCIENCE TECHNIQUES FOR IOT DATA

Data science techniques are crucial in extracting meaningful insights and knowledge from the vast amounts of data generated by IoT devices. These techniques enable organizations to leverage the potential of IoT data for making informed decisions, optimizing processes, and gaining a competitive edge. Tables II to V provides additional details on data science techniques used in IoT.

#### A. Data Preprocessing and Cleaning

Data preprocessing and cleaning are crucial steps in the data science pipeline when dealing with IoT data. Due to the nature of IoT data, which is often generated from diverse sources and in real-time, it is essential to preprocess and clean the data to ensure its quality and usability for further analysis. This involves several techniques to address common challenges associated with IoT data, such as noise, missing values, and inconsistencies.

1) *Noise removal*: IoT data can be susceptible to noise due to various factors, including sensor inaccuracies, communication errors, or environmental interference [39]. Data scientists employ techniques such as smoothing algorithms, filtering, and outlier detection methods to eliminate noise and ensure the accuracy of the data.

2) *Missing data handling*: IoT data streams may encounter missing values due to device failures, network interruptions, or sensor malfunctions [40]. Data scientists utilize imputation methods (e.g., mean imputation, interpolation) or advanced ML techniques to fill in missing data points based on patterns and relationships within the dataset.

3) *Data integration*: IoT applications often involve multiple sensors or devices that generate data in different formats or structures. Data integration techniques combine and merge data from various sources, ensuring consistency and enabling comprehensive analysis [41].

4) *Data transformation*: IoT data may require modification to align with specific analysis requirements or to normalize data across different sensors or devices. Scaling, normalization, and feature engineering are applied to transform the data into a suitable format for subsequent analysis [42].

5) *Data validation and quality assurance*: Data scientists validate IoT data to identify any inconsistencies, errors, or anomalies that may impact the analysis. This involves conducting data quality checks, verifying data integrity, and performing statistical tests to ensure the reliability of the dataset [43].

6) *Time-series analysis*: IoT data often exhibit temporal dependencies and trends. Data scientists leverage time-series analysis techniques to extract meaningful insights from time-

stamped IoT data, such as detecting patterns, forecasting future trends, or identifying anomalies [44].

By applying these data preprocessing and cleaning techniques, data scientists can ensure the quality, reliability, and integrity of IoT data, enabling more accurate and meaningful analysis. These steps lay the foundation for subsequent data science tasks, such as feature selection, model building, and predictive analytics, to derive valuable insights and make informed decisions based on IoT data. Table II summarizes various data preprocessing approaches used in IoT data analysis. Each technique has strengths and shortcomings that researchers and practitioners should consider when preparing IoT data for analysis.

### B. Data Fusion and Integration

Data Fusion and Integration are essential aspects of Data Science Techniques for IoT Data. They involve combining data from various sources and integrating them into a unified dataset for further analysis. Here, we will discuss some commonly used techniques in Data Fusion and Integration for IoT Data:

1) *Sensor data integration*: In IoT systems, data is collected from multiple sensors deployed in different locations. Sensor data integration techniques combine data from various sensors to comprehensively view the environment or system being monitored [45].

2) *Data alignment and synchronization*: IoT devices often have different sampling rates and formats. Data alignment techniques ensure that data from various sources are synchronized and aligned in terms of time and format. This enables accurate analysis and interpretation of the integrated dataset [46].

3) *Data fusion*: Data fusion techniques combine data from multiple sources to derive more accurate and comprehensive insights. This can include techniques like statistical averaging, weighted aggregation, or model-based fusion. Data fusion helps to improve the reliability and accuracy of the integrated dataset [47].

4) *Contextual data integration*: IoT data often includes contextual information such as location, time, and environmental conditions. Contextual data integration techniques aim to incorporate this additional information into the dataset, enabling more profound analysis and correlation with other variables [48].

5) *Semantic data integration*: Semantic data integration techniques focus on incorporating domain-specific knowledge and ontologies to enhance the understanding and interpretation of the integrated dataset. This helps to establish meaningful relationships between different data sources and enables more advanced analytics [49].

By applying these Data Science Techniques for IoT Data Fusion and Integration, organizations can leverage combined data from diverse sources to gain deeper insights, make informed decisions, and derive maximum value from their IoT deployments. Table III provides an overview of data fusion and integration approaches used in IoT data analysis. Each

technique offers unique strengths and may have specific challenges that should be considered when integrating data from multiple sources.

### C. Machine Learning

ML techniques are crucial in analyzing and extracting valuable insights from IoT data. Here, we will discuss some fundamental data science techniques for IoT data that utilize ML:

1) *Anomaly detection*: ML algorithms can detect anomalies in IoT data, which can indicate unusual behavior, faults, or security breaches. By training models on standard data patterns, any deviations from the norm can be identified and flagged for further investigation [50].

2) *Predictive maintenance*: ML models can be employed to predict the maintenance needs of IoT devices and systems. By analyzing historical data, sensor readings, and environmental conditions, predictive maintenance models can anticipate when maintenance or repairs are required, minimizing downtime and optimizing maintenance schedules [51].

3) *Classification and regression*: ML algorithms can be used for classification and regression tasks on IoT data. For example, classification models can classify sensor readings into categories or identify specific events or conditions. Regression models can predict numerical values based on input variables, such as predicting energy consumption based on environmental factors [52].

4) *Clustering and segmentation*: ML clustering algorithms can group similar IoT data instances based on their characteristics or behavior. This can help identify patterns, segment data for targeted analysis, or detect clusters of devices with similar usage patterns [53].

5) *Feature selection and dimensionality reduction*: IoT data can be high-dimensional and contain numerous features. ML techniques like feature selection and dimensionality reduction can identify the most relevant features or transform the data into a lower-dimensional space, improving computational efficiency and model performance [54].

By applying these ML techniques to IoT data, organizations can uncover hidden patterns, make accurate predictions, optimize resource allocation, and gain valuable insights to support decision-making processes. However, it is important to carefully select and train ML models, considering IoT data's specific characteristics and challenges, such as data volume, velocity, variety, and veracity. Table IV provides insights into the various machine learning approaches employed in IoT data analysis. A particular technique for an IoT application should be chosen based on its strengths and limitations.

### D. Anomaly Detection and Outlier Analysis

Anomaly detection and outlier analysis are essential data science techniques used in IoT data to identify unusual patterns, deviations, or outliers that may indicate potential anomalies or anomalies [55]. These techniques are valuable for detecting anomalies in real-time IoT data streams and

addressing security threats, system failures, or abnormal behavior. Anomaly detection involves identifying data instances that deviate significantly from the expected or normal behavior. This can be achieved through various approaches, including statistical methods, ML algorithms, and pattern recognition techniques. The goal is to automatically distinguish between normal and abnormal data instances without prior knowledge of the specific anomalies. In the case of IoT data, anomaly detection can be particularly challenging due to the high volume, velocity, and variety of data generated by IoT devices. Traditional statistical methods, such as mean-based or standard deviation-based approaches, may not be suitable for handling the complexity and dynamics of IoT data. Instead, ML algorithms such as clustering, density-based methods, or ensemble techniques are often employed.

On the other hand, Outlier analysis focuses on identifying data points significantly different from the rest of the dataset. Outliers can arise due to measurement errors, system failures,

or malicious activities [56]. By detecting and analyzing outliers, organizations can gain insights into system vulnerabilities, identify potential risks, and take appropriate actions to mitigate them. Data science techniques for anomaly detection and outlier analysis in IoT data involve several steps. These include data preprocessing, feature engineering, model selection, training, and evaluation. The choice of techniques and algorithms depends on the specific characteristics of the IoT data and the desired level of accuracy and interpretability. Overall, anomaly detection and outlier analysis techniques are essential for ensuring the integrity, security, and reliability of IoT systems. By effectively identifying and responding to anomalies in real-time, organizations can mitigate risks, optimize operations, and enhance the overall performance of their IoT deployments. Table V presents an overview of different anomaly detection and outlier analysis approaches applied to IoT data. Each technique offers distinct strengths and potential challenges, providing researchers with insights into their suitability for specific IoT data scenarios.

TABLE II. DATA PREPROCESSING APPROACHES FOR IOT DATA

Technique	Strengths	Shortcomings
Data cleaning	Improves data quality and accuracy for analysis Removes noise, outliers, and inconsistencies	It may result in data loss if too many data points are removed Manual data cleaning can be time-consuming for large datasets
Missing data handling	Allows for analysis with incomplete data Preserves data integrity and prevents bias	Imputation methods may introduce additional uncertainty Imputed values may not accurately represent the missing data
Data normalization	Enhances data comparability and compatibility Reduces the impact of varying scales and units	Different normalization methods may yield different results Extreme values may distort the normalization process
Feature engineering	Creates informative and relevant features for analysis Capture complex relationships and patterns in the data.	Requires domain expertise to identify meaningful features It may introduce biases if features are not carefully engineered

TABLE III. DATA FUSION AND INTEGRATION APPROACHES FOR IOT DATA

Technique	Strengths	Shortcomings
Data fusion	Integrates data from multiple sources to provide a comprehensive view Enhances data quality and completeness Enables more accurate and holistic analysis	Requires careful handling of data heterogeneity and compatibility Complex integration processes may introduce errors
Data integration	Merges data from different formats, systems, or platforms Enable unified analysis and insights.	Data integration may encounter challenges due to varying data schemas and structures. Requires robust integration mechanisms for real-time or large-scale data
Data synchronization	Ensures consistency and timeliness of data across multiple sources Enables real-time analysis and decision-making	Synchronization mechanisms may introduce latency or data inconsistency Complex synchronization processes may impact system performance

TABLE IV. MACHINE LEARNING APPROACHES FOR IOT DATA

Technique	Strengths	Shortcomings
Supervised learning	Enables accurate predictions and classifications Handles well-labeled and structured IoT data	Requires labeled training data, which can be expensive or time-consuming to obtain Performance may degrade if the model encounters unseen or different data patterns.
Unsupervised learning	Discovers hidden patterns and relationships in IoT data Useful for exploratory analysis and anomaly detection	Interpretation of unsupervised learning results can be challenging Difficult to evaluate the performance objectively without ground truth labels
Reinforcement learning	Learns optimal actions and decision-making strategies based on feedback Suitable for dynamic and interactive IoT systems	It may require significant computational resources and time for training Proper reward design and environment modeling are crucial for effective reinforcement learning in IoT settings

TABLE V. ANOMALY DETECTION AND OUTLIER ANALYSIS APPROACHES FOR IOT DATA

Technique	Strengths	Shortcomings
Statistical methods	Detects deviations from normal patterns in IoT data Relatively interpretable and straightforward	It may not capture complex anomalies or patterns Assumes data distributions and assumptions, which may not hold in all IoT scenarios
Machine learning	Identifies anomalies using advanced pattern recognition algorithms Handles high-dimensional and complex IoT data	Requires labeled or anomalous training data for supervised anomaly detection It may be computationally expensive for real-time or large-scale IoT data analysis.
Time series analysis	Captures temporal dependencies and trends in IoT data Enables forecasting and anomaly detection over time	May struggle with irregular or missing time series data Proper modeling and selection of time series techniques require expertise.

#### IV. DISCUSSION

##### A. Smart Cities and Urban Analytics

Data science plays a crucial role in developing smart cities and urban analytics by harnessing the power of IoT data. Through data science techniques, cities can gather and analyze data from various sources, such as sensors, cameras, and social media, to gain valuable insights and make informed urban planning and management decisions. Data science is applied in smart cities and urban analytics in many ways, including:

1) *Traffic management*: Data science algorithms can process real-time data from traffic sensors and cameras to optimize traffic flow, identify hotspots, and suggest alternative routes to reduce traffic congestion and improve transportation efficiency.

2) *Energy optimization*: By analyzing data from smart meters, energy consumption patterns can be identified, allowing for effective energy management strategies. Data science can help optimize energy distribution, monitor power usage, and identify energy-saving opportunities in buildings and infrastructure.

3) *Waste management*: Data science techniques can analyze data from IoT-enabled waste bins and sensors to optimize waste collection routes, predict bin fill levels, and minimize operational costs. This ensures efficient waste management and contributes to environmental sustainability.

4) *Public safety*: Data science can analyze data from various sources, such as surveillance cameras, social media, and emergency service calls, to detect patterns and trends related to crime, accidents, and emergencies. This enables proactive measures for public safety and emergency response planning.

5) *Urban planning*: By integrating data from multiple sources, including transportation, infrastructure, and social demographics, data science can support urban planners in making informed decisions regarding land use, zoning, and resource allocation. This facilitates the development of sustainable and livable cities.

##### B. Healthcare and Remote Monitoring

Data science has revolutionized the healthcare industry by enabling advanced analytics and insights from IoT data, leading to enhanced healthcare delivery and remote monitoring capabilities. The application of data science in healthcare and remote monitoring offers various benefits, including improved

patient outcomes, personalized treatments, and efficient resource allocation. Data science has various critical applications in healthcare and remote monitoring, which include:

1) *Remote patient monitoring*: Data science techniques analyze data from IoT devices such as wearables, sensors, and mobile apps to monitor patients' vital signs, activity levels, and medication adherence remotely. This enables healthcare providers to detect anomalies, track patient progress, and intervene promptly if necessary.

2) *Predictive analytics for disease prevention*: By analyzing large volumes of healthcare data, including patient records, genetic information, and environmental factors, data science can identify patterns and risk factors for diseases. This helps in early detection, prevention, and personalized treatment planning.

3) *Real-time health monitoring*: Data science algorithms process real-time sensor data to continuously monitor patients' health conditions. This enables early detection of critical events, such as cardiac abnormalities or falls, and alerts healthcare providers for immediate intervention.

4) *Healthcare resource optimization*: Data science techniques optimize healthcare resource allocation by analyzing patient flow, resource utilization, and demand forecasting data. This helps healthcare organizations streamline operations, reduce waiting times, and allocate resources efficiently.

5) *Personalized medicine*: Data science enables the analysis of large-scale genomic and patient data to develop customized treatment plans based on individual characteristics, genetic markers, and treatment response patterns. This promotes precision medicine and improves patient outcomes.

##### C. Agriculture and Precision Farming

Data science has emerged as a valuable tool in agriculture and precision farming, enabling farmers to optimize crop production, improve resource management, and make data-driven decisions. The application of data science in agriculture leverages IoT devices, sensors, and data analytics to monitor and analyze various parameters related to soil, weather, crops, and farm operations. Agriculture and precision farming benefit from data science in multiple ways, including:

1) *Crop yield prediction*: Data science techniques analyze historical and real-time data on soil conditions, weather

patterns, crop health, and farming practices to predict crop yields. This helps farmers make informed decisions regarding planting schedules, irrigation, fertilization, and pest management.

2) *Precision irrigation*: Data science algorithms process data from soil moisture sensors, weather forecasts, and crop water requirements to optimize irrigation practices. This ensures that crops receive the right amount of water at the right time, minimizing water wastage and reducing the risk of water stress.

3) *Disease and pest management*: Data science models analyze data from IoT devices, such as insect traps, disease sensors, and satellite imagery, to detect and monitor pests, diseases, and weed infestations. This enables early intervention and targeted treatment, reducing the use of pesticides and minimizing crop losses.

4) *Nutrient management*: Data science techniques analyze soil composition data, crop nutrient requirements, and historical yield data to optimize fertilization practices. This ensures crops receive the necessary nutrients correctly, promoting healthy growth and maximizing yield.

5) *Farm equipment optimization*: Data science algorithms analyze data from IoT-enabled farm equipment, such as tractors and harvesters, to optimize their usage, fuel consumption, and maintenance schedules. This helps farmers improve operational efficiency, reduce costs, and prolong the lifespan of equipment.

#### D. Industrial IoT and Predictive Maintenance

Data science plays a crucial role in Industrial IoT (IIoT) by enabling predictive maintenance, optimizing operations, and improving overall efficiency in industrial settings. Applying data science techniques in IIoT allows for real-time monitoring, analysis, and prediction of equipment health and performance. IIoT and predictive maintenance are two areas where data science finds crucial applications, such as:

1) *Predictive maintenance*: Data science models analyze data from IoT sensors, equipment logs, and historical maintenance records to proactively predict equipment failures and schedule maintenance activities. By identifying potential issues before they occur, companies can avoid costly unplanned downtime and optimize maintenance schedules, leading to increased productivity and reduced maintenance costs.

2) *Asset performance optimization*: Data science techniques analyze sensor data and operational parameters to optimize asset performance and efficiency. By monitoring key performance indicators and analyzing historical data, companies can identify areas for improvement, optimize energy consumption, and enhance overall equipment effectiveness (OEE).

3) *Quality control and defect detection*: Data science algorithms analyze sensor data and production metrics to detect anomalies, identify patterns, and ensure product quality. By monitoring and analyzing real-time data, companies can

identify quality issues early, reduce defects, and improve overall product quality and customer satisfaction.

4) *Supply chain optimization*: Data science techniques can analyze data from IoT devices, inventory records, and transportation systems to optimize supply chain operations. By predicting demand, optimizing inventory levels, and improving logistics, companies can streamline their supply chain processes, reduce costs, and improve customer service.

5) *Process optimization*: To optimize industrial processes, data science models analyze sensor data, production parameters, and historical data. By identifying inefficiencies, bottlenecks, and areas for improvement, companies can optimize process parameters, reduce waste, and improve overall productivity.

#### V. OPEN RESEARCH CHALLENGES AND FUTURE DIRECTIONS

- **Privacy and security in IoT data analytics**: As the IoT expands, ensuring privacy and security becomes a paramount concern. Researchers must address the challenges of securing IoT data throughout its lifecycle, including data collection, storage, transmission, and analysis. Developing robust encryption techniques, access control mechanisms, and secure data-sharing protocols will be crucial. Additionally, exploring privacy-preserving data analytics methods, such as federated learning or differential privacy, can help protect sensitive IoT data while extracting meaningful insights.
- **Interpretable and explainable ML models**: As ML algorithms play a crucial role in IoT data analytics, it is essential to develop interpretable and explainable models. The transparency of models becomes increasingly important in critical domains such as healthcare, where trust and accountability are paramount. Researchers should develop techniques to enhance model interpretability, including feature importance analysis, rule-based models, and model-agnostic explanation methods. This will enable users to understand the reasoning behind the predictions and decisions made by the models.
- **Ethical considerations in data science for IoT**: Integrating data science and IoT raises ethical concerns that must be addressed. Researchers must explore ethical frameworks and guidelines for IoT data collection, usage, and governance. Ensuring informed consent, anonymization, and fair and unbiased data analysis are vital challenges. Ethical decision-making frameworks, transparency in data handling practices, and guidelines for responsible data usage can help address these concerns and promote ethical practices in data science for IoT.
- **Trust and reliability in IoT data analysis**: Building trust in IoT data analysis is crucial for its adoption. Researchers must focus on data quality assurance, integrity verification, and algorithmic fairness in IoT data analytics. Exploring methods for data validation,

anomaly detection, and bias mitigation will help improve the reliability and trustworthiness of IoT data analysis results. Additionally, developing mechanisms to assess and quantify the trustworthiness of IoT data sources and algorithms will contribute to more reliable and robust decision-making processes.

- Scalability and efficiency: As the scale and complexity of IoT systems continue to grow, there is a need for scalable and efficient data science techniques. Researchers should focus on developing algorithms and frameworks that can handle the massive volume of data generated by IoT devices and efficiently process it in real-time. Techniques such as distributed computing, edge computing, and stream processing can be vital in addressing scalability challenges.
- Real-time and stream analytics: The time-sensitive nature of IoT data requires real-time and stream analytics capabilities. Researchers must focus on developing algorithms and frameworks to process streaming data in real-time and extract meaningful insights. Complex event processing, predictive analytics, and online learning can enable real-time decision-making and proactive responses based on IoT data.
- Edge and fog computing: The advent of edge and fog computing brings new opportunities and challenges to data science for IoT. Researchers should explore how data science techniques can be integrated with edge and fog computing architectures to enable real-time analytics and decision-making at the network edge. Developing efficient data processing and analytics algorithms tailored explicitly for edge and fog environments will be crucial.
- Federated learning: As IoT devices are distributed across various networks and locations, federated learning presents an opportunity to train ML models directly on edge devices while preserving data privacy. Researchers should explore efficient and secure federated learning techniques in IoT environments, considering limited computational resources, heterogeneous data sources, and communication constraints.
- Context-aware analytics: IoT data is inherently contextual, capturing information about the physical environment, user behavior, and situational context. Incorporating context awareness into data science models and algorithms can lead to more accurate and personalized insights. Researchers should investigate methods for context-aware analytics, including techniques for context acquisition, context representation, and context-aware modeling and prediction.
- Integration of domain knowledge: IoT data often carries domain-specific characteristics and semantics. Incorporating domain knowledge into data science models can enhance their performance and interpretability. Researchers should focus on developing

techniques for integrating domain knowledge into IoT data analytics, leveraging ontologies, expert systems, and domain-specific feature engineering approaches.

- Adaptive and self-learning systems: IoT environments are dynamic and evolve over time. Developing adaptive and self-learning systems for IoT data science can enable models to continuously learn and adapt to changing conditions. Researchers should explore online, reinforcement, and transfer learning to build intelligent systems that adapt to evolving IoT data streams and environments.

## VI. CONCLUSION

The field of data science for the IoT holds immense potential in unlocking valuable insights and enabling transformative applications. This paper has provided a comprehensive overview of the key concepts, challenges, and techniques in this emerging field. We have explored the diverse applications of Data Science in IoT across various domains, including smart cities, healthcare, agriculture, and industrial IoT. These applications have showcased the ability of Data Science to revolutionize industries, optimize processes, and improve decision-making. Throughout the discussion, we have highlighted the significant challenges associated with IoT data, such as volume, velocity, variety, and veracity. We have examined the techniques and methodologies to address these challenges, including data preprocessing and cleaning, data fusion and integration, ML, anomaly detection, and outlier analysis. These techniques provide valuable insights from the vast and heterogeneous IoT data, enabling organizations to make data-driven decisions and derive actionable intelligence.

Furthermore, we have delved into the importance of scalable data processing, distributed computing frameworks, and edge computing for handling massive amounts of IoT data and facilitating real-time analytics. We have discussed stream processing and real-time analytics as essential components for processing data streams and extracting immediate insights from dynamic IoT environments. While Data Science for IoT presents numerous opportunities, open research challenges, and future directions still need to be addressed. These include privacy and security considerations in IoT data analytics, developing interpretable and explainable ML models, ethical considerations in data science for IoT, and ensuring trust and reliability in IoT data analysis. These areas require further exploration and innovation to fully harness the potential of IoT data while ensuring responsible and ethical practices.

## REFERENCES

- [1] M. Mohseni, F. Amirghafouri, and B. Pourghebleh, "CEDAR: A cluster-based energy-aware data aggregation routing protocol in the internet of things using capuchin search algorithm and fuzzy logic," *Peer-to-Peer Networking and Applications*, pp. 1-21, 2022.
- [2] F. Kamalov, B. Pourghebleh, M. Gheisari, Y. Liu, and S. Moussa, "Internet of Medical Things Privacy and Security: Challenges, Solutions, and Future Trends from a New Perspective," *Sustainability*, vol. 15, no. 4, p. 3317, 2023.
- [3] 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.

- [4] 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.
- [5] 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.
- [6] 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.
- [7] H. Kosarirad, M. Ghasempour Nejati, A. Saffari, M. Khishe, and M. Mohammadi, "Feature Selection and Training Multilayer Perceptron Neural Networks Using Grasshopper Optimization Algorithm for Design Optimal Classifier of Big Data Sonar," *Journal of Sensors*, vol. 2022, 2022.
- [8] R. Soleimani and E. Lobaton, "Enhancing Inference on Physiological and Kinematic Periodic Signals via Phase-Based Interpretability and Multi-Task Learning," *Information*, vol. 13, no. 7, p. 326, 2022.
- [9] B. M. Jafari, M. Zhao, and A. Jafari, "Rumi: An Intelligent Agent Enhancing Learning Management Systems Using Machine Learning Techniques," *Journal of Software Engineering and Applications*, vol. 15, no. 9, pp. 325-343, 2022.
- [10] M. Sarbaz, M. Manthouri, and I. Zamani, "Rough neural network and adaptive feedback linearization control based on Lyapunov function," in *2021 7th International Conference on Control, Instrumentation and Automation (ICCIA)*, 2021: IEEE, pp. 1-5.
- [11] C. Han and X. Fu, "Challenge and Opportunity: Deep Learning-Based Stock Price Prediction by Using Bi-Directional LSTM Model," *Frontiers in Business, Economics and Management*, vol. 8, no. 2, pp. 51-54, 2023.
- [12] S. Saeidi, S. Enjedani, E. Alvandi Behineh, K. Tehranian, and S. Jazayerifar, "Factors Affecting Public Transportation Use during Pandemic: An Integrated Approach of Technology Acceptance Model and Theory of Planned Behavior," *Tehnički glasnik*, vol. 18, pp. 1-12, 09/01 2023, doi: 10.31803/tg-20230601145322.
- [13] J. Bhatia et al., "An overview of fog data analytics for IoT applications," *Sensors*, vol. 23, no. 1, p. 199, 2022.
- [14] S. Ayyaz and K. Alpay, "Predictive maintenance system for production lines in manufacturing: A machine learning approach using IoT data in real-time," *Expert Systems with Applications*, vol. 173, p. 114598, 2021.
- [15] E. S. Pramukantoro, D. P. Kartikasari, and R. A. Siregar, "Performance evaluation of MongoDB, cassandra, and HBase for heterogenous IoT data storage," in *2019 2nd International Conference on Applied Information Technology and Innovation (ICAITI)*, 2019: IEEE, pp. 203-206.
- [16] Goknil et al., "A Systematic Review of Data Quality in CPS and IoT for Industry 4.0," *ACM Computing Surveys*, 2023.
- [17] V. C. Farias da Costa, L. Oliveira, and J. de Souza, "Internet of everything (IoE) taxonomies: A survey and a novel knowledge-based taxonomy," *Sensors*, vol. 21, no. 2, p. 568, 2021.
- [18] S. R. Poojara, C. K. Dehury, P. Jakovits, and S. N. Srirama, "Serverless data pipeline approaches for IoT data in fog and cloud computing," *Future Generation Computer Systems*, vol. 130, pp. 91-105, 2022.
- [19] K. Y. Lee, M. Seo, R. Lee, M. Park, and S.-H. Lee, "Efficient processing of spatio-temporal joins on IoT data," *IEEE Access*, vol. 8, pp. 108371-108386, 2020.
- [20] S. Akiyoshi, Y. Taenaka, K. Tsukamoto, and M. Lee, "Loose Matching Approach Considering the Time Constraint for Spatio-Temporal Content Discovery," in *Advances in Intelligent Networking and Collaborative Systems: The 13th International Conference on Intelligent Networking and Collaborative Systems (INCoS-2021)* 13, 2022: Springer, pp. 295-306.
- [21] Y. Sasaki, "A survey on IoT big data analytic systems: Current and future," *IEEE Internet of Things Journal*, vol. 9, no. 2, pp. 1024-1036, 2021.
- [22] K. Ahmad, M. Maabreh, M. Ghaly, K. Khan, J. Qadir, and A. Al-Fuqaha, "Developing future human-centered smart cities: Critical analysis of smart city security, Data management, and Ethical challenges," *Computer Science Review*, vol. 43, p. 100452, 2022.
- [23] S. Namani and B. Gonen, "Smart agriculture based on IoT and cloud computing," in *2020 3rd International Conference on Information and Computer Technologies (ICICT)*, 2020: IEEE, pp. 553-556.
- [24] Wang, J. Xie, Y. Wan, G. A. Guijarro Reyes, and L. R. Garcia Carrillo, "3-d trajectory modeling for unmanned aerial vehicles," in *AIAA Scitech 2019 Forum*, 2019, p. 1061.
- [25] N. S. Baqer, H. Mohammed, and A. Albahri, "Development of a real-time monitoring and detection indoor air quality system for intensive care unit and emergency department," *Signa Vitae*, vol. 19, no. 1, 2023.
- [26] 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, doi: <https://doi.org/10.1016/j.suscom.2022.100822>.
- [27] Razmjoo, P. A. Østergaard, M. Denai, M. M. Nezhad, and S. Mirjalili, "Effective policies to overcome barriers in the development of smart cities," *Energy Research & Social Science*, vol. 79, p. 102175, 2021.
- [28] J. Azar, A. Makhoul, M. Barhamgi, and R. Couturier, "An energy efficient IoT data compression approach for edge machine learning," *Future Generation Computer Systems*, vol. 96, pp. 168-175, 2019.
- [29] L. Babangida, T. Perumal, N. Mustapha, and R. Yaakob, "Internet of things (IoT) based activity recognition strategies in smart homes: A review," *IEEE Sensors Journal*, vol. 22, no. 9, pp. 8327-8336, 2022.
- [30] Pourghebleh, N. Hekmati, Z. Davoudnia, and M. Sadeghi, "A roadmap towards energy-efficient data fusion methods in the Internet of Things," *Concurrency and Computation: Practice and Experience*, p. e6959, 2022.
- [31] V. Hayyolalam, B. Pourghebleh, 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.
- [32] W. Li, M. Batty, and M. F. Goodchild, "Real-time GIS for smart cities," vol. 34, ed: Taylor & Francis, 2020, pp. 311-324.
- [33] B. Pourghebleh and V. Hayyolalam, "A comprehensive and systematic review of the load balancing mechanisms in the Internet of Things," *Cluster Computing*, pp. 1-21, 2019.
- [34] H.-C. Lu, F. Hwang, and Y.-H. Huang, "Parallel and distributed architecture of genetic algorithm on Apache Hadoop and Spark," *Applied Soft Computing*, vol. 95, p. 106497, 2020.
- [35] B. Pourghebleh, K. Wakil, and N. J. Navimipour, "A comprehensive study on the trust management techniques in the Internet of Things," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 9326-9337, 2019.
- [36] X. Sáez-de-Cámara, J. L. Flores, C. Arellano, A. Urbietta, and U. Zurutuza, "Clustered Federated Learning Architecture for Network Anomaly Detection in Large Scale Heterogeneous IoT Networks," *Computers & Security*, p. 103299, 2023.
- [37] 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.
- [38] R. Saura, D. Ribeiro-Soriano, and D. Palacios-Marqués, "Assessing behavioral data science privacy issues in government artificial intelligence deployment," *Government Information Quarterly*, vol. 39, no. 4, p. 101679, 2022.
- [39] X.-B. Jin et al., "Deep-learning temporal predictor via bidirectional self-attentive encoder-decoder framework for IOT-based environmental sensing in intelligent greenhouse," *Agriculture*, vol. 11, no. 8, p. 802, 2021.
- [40] R. Krishnamurthi, A. Kumar, D. Gopinathan, A. Nayyar, and B. Qureshi, "An overview of IoT sensor data processing, fusion, and analysis techniques," *Sensors*, vol. 20, no. 21, p. 6076, 2020.
- [41] Chen, L. Ramanathan, and M. Alazab, "Holistic big data integrated artificial intelligent modeling to improve privacy and security in data management of smart cities," *Microprocessors and Microsystems*, vol. 81, p. 103722, 2021.
- [42] B. Pourghebleh, V. Hayyolalam, and A. A. Anvigh, "Service discovery in the Internet of Things: review of current trends and research challenges," *Wireless Networks*, vol. 26, no. 7, pp. 5371-5391, 2020.

- [43] H. Foidl and M. Felderer, "An approach for assessing industrial IoT data sources to determine their data trustworthiness," *Internet of Things*, vol. 22, p. 100735, 2023.
- [44] B. Zhu et al., "IoT equipment monitoring system based on C5. 0 decision tree and time-series analysis," *IEEE Access*, vol. 10, pp. 36637-36648, 2021.
- [45] S. Balakrishna, M. Thirumaran, and V. K. Solanki, "IoT sensor data integration in healthcare using semantics and machine learning approaches," *A handbook of internet of things in biomedical and cyber physical system*, pp. 275-300, 2020.
- [46] S. Sguazza et al., "Sensor data synchronization in a IoT environment for infants motricity measurement," in *IoT Technologies for HealthCare: 6th EAI International Conference, HealthyIoT 2019, Braga, Portugal, December 4–6, 2019, Proceedings 6, 2020: Springer*, pp. 3-21.
- [47] X. Huang, Y. Liu, L. Huang, E. Onstein, and C. Merschbrock, "BIM and IoT data fusion: The data process model perspective," *Automation in Construction*, vol. 149, p. 104792, 2023.
- [48] S. Dalenogare, M.-A. Le Dain, G. B. Benitez, N. F. Ayala, and A. G. Frank, "Multichannel digital service delivery and service ecosystems: The role of data integration within Smart Product-Service Systems," *Technological Forecasting and Social Change*, vol. 183, p. 121894, 2022.
- [49] Teniente, "Iot semantic data integration through ontologies," in *2022 IEEE International Conference on Services Computing (SCC), 2022: IEEE*, pp. 357-358.
- [50] A. Cook, G. Misrlı, and Z. Fan, "Anomaly detection for IoT time-series data: A survey," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 6481-6494, 2019.
- [51] J. C. Cheng, W. Chen, K. Chen, and Q. Wang, "Data-driven predictive maintenance planning framework for MEP components based on BIM and IoT using machine learning algorithms," *Automation in Construction*, vol. 112, p. 103087, 2020.
- [52] P. M. Shakeel, S. Baskar, H. Fouad, G. Manogaran, V. Saravanan, and Q. Xin, "Creating collision-free communication in IoT with 6G using multiple machine access learning collision avoidance protocol," *Mobile Networks and Applications*, vol. 26, pp. 969-980, 2021.
- [53] J. Li, W. Cui, A. Zeng, Y. Xie, and S. Yang, "Clinical Analysis of Medical IoT and Acute Cerebral Infarction Based on Image Recognition," *Mobile Information Systems*, vol. 2022, 2022.
- [54] R. Samdekar, S. Ghosh, and K. Srinivas, "Efficiency enhancement of intrusion detection in iot based on machine learning through bioinspire," in *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), 2021: IEEE*, pp. 383-387.
- [55] Ullah and Q. H. Mahmoud, "Design and development of RNN anomaly detection model for IoT networks," *IEEE Access*, vol. 10, pp. 62722-62750, 2022.
- [56] Kumar, M. Yadav, and A. Chauhan, "Outlier Analysis Based Intrusion Detection for IoT," in *2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), 2021: IEEE*, pp. 1341-1348.



# A Fruit Ripeness Detection Method using Adapted Deep Learning-based Approach

Weiwei Zhang\*

Henan Polytechnic Institute, Nanyang, Henan 473000, China

**Abstract**—Fruit ripeness detection plays a crucial role in precise agriculture, enabling optimal harvesting and post-harvest handling. Various methods have been investigated in the literature for fruit ripeness detection in vision-based systems, with deep learning approaches demonstrating superior accuracy compared to other approaches. However, the current research challenge lies in achieving high accuracy rates in deep learning-based fruit ripeness detection. In this study proposes a method based on the YOLOv8 algorithm to address this challenge. The proposed method involves generating a model using a custom dataset and conducting training, validation, and testing processes. Experimental results and performance evaluation demonstrate the effectiveness of the proposed method in achieving accurate fruit ripeness detection. The proposed method surpasses existing approaches through extensive experiments and performance analysis, providing a reliable solution for fruit ripeness detection in precise agriculture.

**Keywords**—Fruit ripeness detection; precise agriculture; deep learning; vision system; YOLOv8

## I. INTRODUCTION

Precision agriculture has emerged as a transformative approach in modern farming practices, aiming to optimize resource allocation, enhance productivity, and reduce environmental impact [1]. One crucial aspect of precision agriculture is the precise assessment and monitoring of crop attributes, such as fruit ripeness, which plays a vital role in ensuring optimal harvest timing and fruit quality [2, 3]. Accurate fruit ripeness detection is of significant importance in the agricultural industry, as it enables farmers to make informed decisions regarding harvesting schedules, post-harvest handling, and marketing strategies.

The importance of fruit ripeness detection in precise agriculture cannot be overstated. Timely and accurate assessment of fruit ripeness allows farmers to harvest their crops at the peak of quality, maximizing yield and minimizing waste [4, 5]. Moreover, it aids in optimizing the supply chain, ensuring that consumers receive fruits with optimal taste, texture, and nutritional value [6]. Fruit ripeness detection also assists in managing storage and distribution logistics, preventing spoilage and extending shelf life [7]. Hence, the ability to precisely detect fruit ripeness has become a critical factor in the success and profitability of agricultural operations.

Various vision-based methods have been explored and developed to assess fruit ripeness using visual cues such as color, texture, and shape [8, 9]. These methods leverage image processing algorithms and machine learning models to extract meaningful features and classify fruits based on their ripeness

levels [10, 20]. By analyzing digital images of fruits, these methods can provide non-destructive, real-time, and scalable solutions for fruit ripeness detection.

Previous studies have shown a growing interest in deep learning-based methods for fruit ripeness detection due to their ability to automatically learn and extract complex features from large-scale datasets [10]. Deep learning models, such as Convolutional Neural Networks (CNNs), have demonstrated remarkable performance in various computer vision tasks, including object recognition and image classification and other related applications [11, 12, 19]. Researchers have adopted deep learning approaches to develop robust and accurate models for fruit ripeness detection, overcoming some of the limitations of traditional image processing techniques.

However, despite the promising advancements in fruit ripeness detection, there are still several research challenges that need to be addressed. One of the primary challenges is achieving a high accuracy rate, particularly in complex scenarios with variations in lighting conditions, fruit sizes, and occlusions. Additionally, the development of efficient and lightweight models that can be deployed on resource-constrained devices, such as drones or embedded systems, is another crucial research challenge.

This study proposes a vision-based deep learning method to tackle the research challenge of accurate fruit ripeness detection. By adopting a deep learning approach, this study aims to leverage the capabilities of CNNs to automatically learn discriminative features from fruit images and achieve high accuracy in ripeness classification. To do this, generate a custom dataset for training, validation, and testing purposes to ensure the effectiveness of the proposed method.

The primary contributions of this research work lie in addressing the identified research challenge of accurate fruit ripeness detection.

- Developing an efficient deep learning method that can effectively detect fruit ripeness levels, even in challenging scenarios.
- Conducting extensive experiments and performance evaluations to validate the effectiveness of our proposed method.
- Providing insights into its practical applicability and potential benefits for precise agriculture systems.

## II. REVIEW OF PREVIOUS STUDIES

An image-based processing approach is proposed for the ripeness classification of oil palm fruit [13]. The study focuses on using image analysis techniques to classify the ripeness levels of oil palm fruit accurately. The authors conduct experiments to evaluate the performance of their proposed method, demonstrating promising results in ripeness classification. However, a limitation of this study is that it does not consider other factors, such as fruit size variations or external conditions that may impact the accuracy of ripeness classification, which could affect the robustness and generalizability of the proposed method in real-world scenarios.

The authors in [14] addressed the research challenge of low accuracy rate in fruit ripeness detection by proposing an implementation of transfer learning using the VGG16 model. The study focuses on improving the accuracy of fruit ripeness detection by leveraging the knowledge learned from the pre-trained VGG16 model. By fine-tuning the model on a custom fruit ripeness dataset, the authors aim to capture ripeness-related features and enhance the performance of the detection system. Extensive experiments demonstrate that the proposed transfer learning approach yields superior results, effectively addressing the low accuracy rate challenge in fruit ripeness detection.

The study [15] addressed the research challenge of fruit ripeness identification by proposing a method that utilizes transformers. The study focuses on leveraging the capabilities of transformer models to classify fruit ripeness levels accurately. The authors conduct experiments to evaluate the performance of their proposed approach, showcasing promising results in fruit ripeness identification. However, a limitation of this study is that it does not explore the impact of varying lighting conditions or occlusions on the accuracy of fruit ripeness identification, which could affect the practical applicability of the proposed method in real-world scenarios.

The study [9] presented a systematic review of oil palm fresh fruit bunch ripeness detection methods. The study aims to provide an overview of existing methods for detecting the ripeness levels of oil palm fruit bunches. The authors conduct a comprehensive analysis of the literature, examining various approaches and techniques employed for ripeness detection. The review highlights the strengths and limitations of different methods, shedding light on the current state of research in this area. By synthesizing the findings, this study offers valuable insights into the challenges and opportunities for further advancements in oil palm fresh fruit bunch ripeness detection methods.

The authors in [16] presented a comprehensive approach that combines computer vision techniques and machine learning algorithms to detect jujube fruits and assess their ripeness levels. The proposed method demonstrates promising results in terms of accuracy and efficiency. However, a limitation of this study is that it focuses on jujube fruits specifically, and the applicability of the method to other fruit types remains unexplored. Further research is needed to

evaluate its effectiveness across different fruit varieties, addressing the generalizability of the proposed method.

## III. MATERIAL AND METHODS

### A. COCO Dataset

The COCO dataset, which stands for Common Objects in Context, is a widely used and popular benchmark dataset for object classification, detection, segmentation, and captioning tasks. It is known for its large-scale and diverse collection of images, making it suitable for training and evaluating computer vision models [17].

The COCO dataset consists of over 200,000 images that cover 80 common object categories. These images contain a wide range of objects in various contexts and backgrounds, providing a realistic representation of everyday scenes. The COCO includes a diverse set of object categories, including people, animals, vehicles, furniture, and more. It captures a broad range of object appearances, poses, and scales. Fig. 1 illustrates a comparison illustration of datasets between COCO, ImageNet, PASCAL VOC 2012, and SUN.

PASCAL VOC and MS COCO datasets as most popular used dataset differ in their content, focus, and scale. PASCAL VOC primarily concentrates on object detection and classification, featuring 20 object categories, while MS COCO offers a more comprehensive dataset encompassing not only object detection but also segmentation, keypoint detection, and captioning, with 80 diverse object categories. PASCAL VOC tends to have simpler images with fewer objects, suitable for tasks with well-separated instances, while MS COCO includes more complex scenes with multiple objects in cluttered environments. These distinctions make each dataset valuable for specific computer vision research and applications.

Fig. 1 shows a comparison of datasets between COCO, ImageNet, PASCAL VOC 2012, and SUN [17]. Fig. 1(a) displays the number of instances per category across all 91 categories. Additionally, Fig. 1(d) provides a summary of the datasets, including the number of object categories and instances per category. Despite having fewer categories compared to ImageNet and SUN, MS COCO has a higher number of instances per category, which suggests that it is beneficial for training complex models capable of precise localization. Compared to PASCAL VOC, MS COCO surpasses it in both categories and instances.

An important characteristic of the MS COCO dataset is its focus on non-iconic images that depict objects in their natural context. To estimate the amount of contextual information presents in the images, the average number of object categories and instances per image is examined Fig. 1(b) and Fig.1(c). In MS COCO, each image average contains 3.5 categories and 7.7 instances. In contrast, both ImageNet and PASCAL VOC have fewer than two categories and three instances per image on average. Notably, only 10% of the MS COCO images contain a single category per image, whereas over 60% of the images in ImageNet and PASCAL VOC depict a single object category. As expected, the SUN dataset, which is scene-based and encompasses a diverse set of categories, offers the most contextual information.

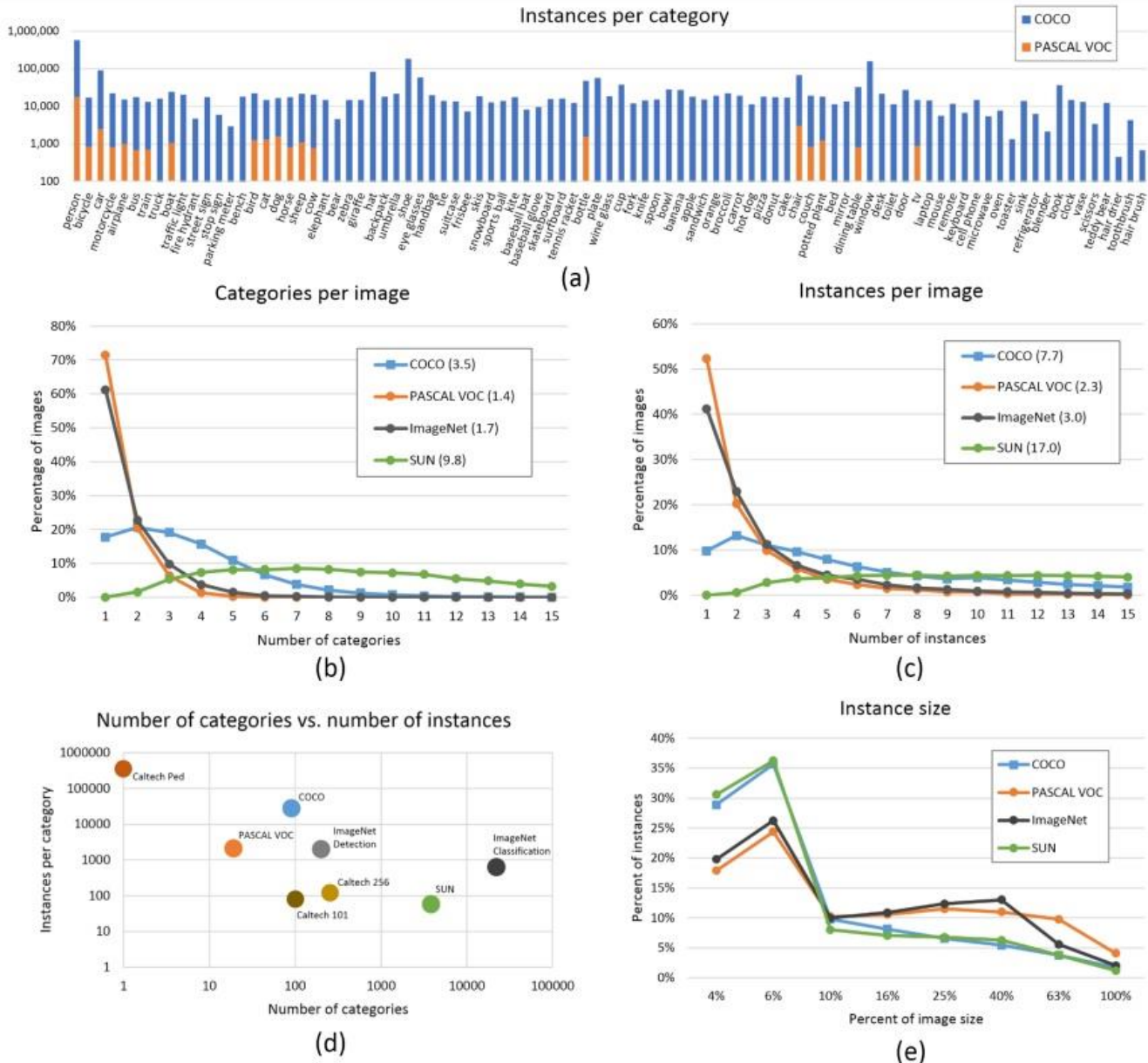


Fig. 1. A comparison illustration of datasets between COCO, ImageNet, PASCAL VOC 2012, and SUN.

### B. YOLOv8 Model

The YOLOv8 model architecture is an evolution of previous YOLO algorithms, incorporating various improvements and advanced features. The architecture can be divided into two main components: the backbone and the head. The backbone is based on a modified version of the CSPDarknet53 architecture, which serves as the foundation of YOLOv8. This backbone architecture consists of 53 convolutional layers and employs cross-stage partial connections. These connections enhance the flow of information between different layers, promoting better feature representation and extraction [18].

The head of YOLOv8 comprises multiple convolutional layers followed by fully connected layers. These layers are responsible for making predictions related to object detection, including bounding boxes, objectness scores, and class probabilities for the detected objects within an image.

The YOLOv8 introduces multi-scaled object detection capabilities. To achieve this, the model utilizes a feature pyramid network. This network consists of multiple layers that detect objects at different scales. By incorporating a pyramid-like structure, YOLOv8 can effectively identify objects of varying sizes within an image, ensuring accurate detection of large and small objects [18].

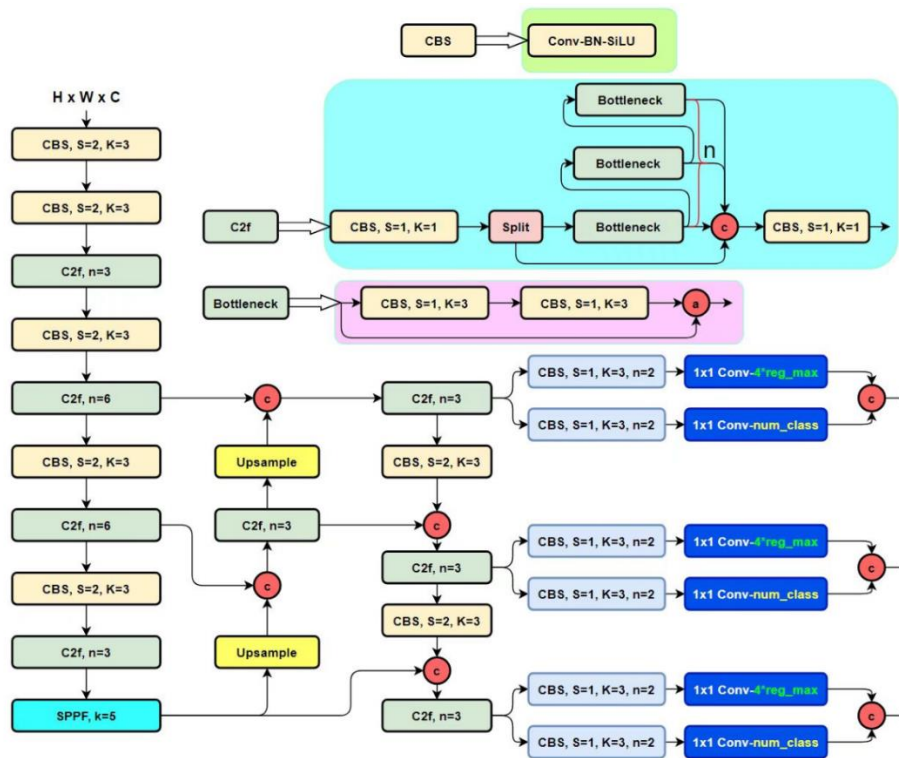


Fig. 2. YOLOv8 architecture [18].

As shown in Fig. 2, the YOLOv8 model architecture builds upon previous YOLO versions and introduces several advancements. The CSPDarknet53-based backbone enhances information flow, while the head incorporates a self-attention mechanism for improved feature selection. Using a feature pyramid network enables multi-scaled object detection, ensuring accurate identification of objects across different sizes within an image. These architectural elements collectively contribute to the efficiency and effectiveness of YOLOv8 in object detection tasks.

### C. Fruit Ripeness Detection Model

This study adopted a YOLOv8 model for fruit ripeness detection. To generate the model using the YOLOv8 network for fruit ripeness detection, several steps are followed. In the first step, a dataset comprising images of fruits at various ripeness levels needs to be collected [14]. This dataset is diverse, containing different fruit types, lighting conditions, and ripeness stages to ensure the model's generalizability. Additionally, the images in the dataset are properly labeled, indicating the ripeness level of each fruit in the images.

Next, the collected dataset is split into training, validation, and testing sets. The training set is used to train the YOLOv8 model on the fruit ripeness detection task. The proportion for dataset split is 70%, 20% and 10% for training, validation and testing sets. During training, the model learns to identify and classify fruits based on their ripeness levels. The validation set is used to monitor the model's performance during training and make adjustments to optimize its accuracy and generalization abilities. Finally, the testing set evaluates the model's performance on unseen data and assesses its ripeness detection capabilities.

Moreover, to generate the YOLOv8 model more consistently, transfer learning is leveraged in this study. Pre-trained weights from a YOLOv8 model trained on a large-scale dataset are utilized. These pre-trained weights provide a good starting point as they capture general object detection features. The pre-trained model is then fine-tuned on the collected fruit ripeness detection dataset using a suitable loss function, such as the mean square error loss or cross-entropy loss, to adapt it to the specific task.

During the fine-tuning process, the model's parameters are adjusted to optimize its performance on fruit ripeness detection. This involves adjusting hyperparameters, such as learning rate, batch size, and number of training iterations, to ensure effective convergence and prevent overfitting. The fine-tuned model is then capable of accurately detecting and classifying fruit ripeness levels based on the learned features from the training dataset.

## IV. RESULTS AND DISCUSSION

### A. Comparison of YOLO Models

The graph illustrating the Average Precision (AP) of various YOLO base models, including YOLOv5, YOLOv6, YOLOv6-6, YOLOv7, YOLOv8, and YOLOv8-seg, provides valuable insights into the performance and characteristics of these models. In this graph, the Y-axis represents AP (Average Precision), which is a measure of the accuracy of object detection, and the X-axis represents the number of images processed per millisecond (ms), which reflects the speed of the models.

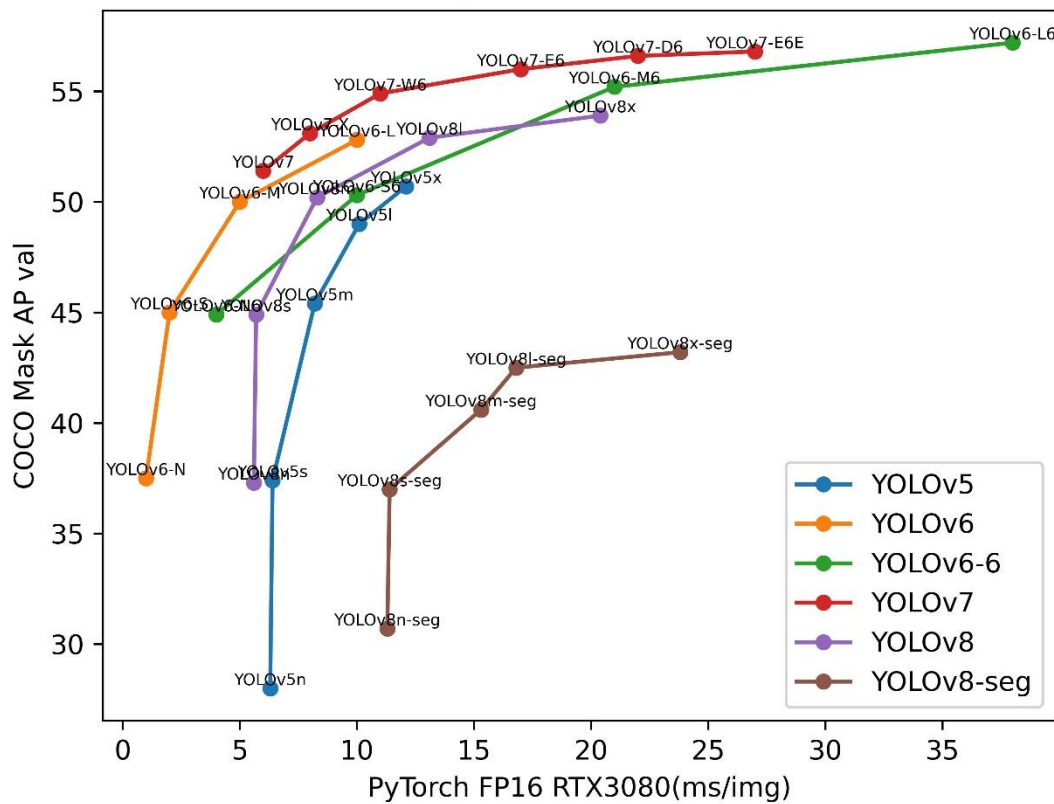


Fig. 3. Performance comparison of Yolo-based models [19].

As illustrated in Fig. 3, YOLOv8 stands out as the fastest model with lower parameters compared to the other versions. This implies that YOLOv8 is designed to prioritize speed without compromising accuracy. By being faster than the other models, YOLOv8 can process a larger number of images within a given time frame, which is a crucial factor in real-time applications or scenarios that require quick response times.

Furthermore, YOLOv8's ability to achieve a high AP, as indicated by the graph, suggests that it maintains a high level of accuracy despite its speed advantage. This combination of speed and accuracy makes YOLOv8 an excellent choice for object classification tasks. Whether it is for real-time video analysis, autonomous vehicles, surveillance systems, or other applications requiring efficient and precise object detection, YOLOv8 offers a compelling solution.

Additionally, the graph shows that YOLOv8 has lower parameters compared to the other models. Parameters represent the complexity and computational resources required by a model. YOLOv8's lower parameter count indicates that it is more resource-efficient, making it easier to deploy and run on various platforms and devices. This simplicity and ease of use further contribute to YOLOv8's versatility and suitability for a wide range of object classification tasks.

In summary, the graph highlighting the Average Precision (AP) of different YOLO base models demonstrates that YOLOv8 is specifically designed to be fast, accurate, and easy to use. Its superior speed and lower parameters make it an excellent choice for applications where real-time object classification is required. YOLOv8's ability to deliver high

accuracy ensures reliable results, and its resource efficiency enhances its usability across various platforms.

### B. Experimental Results

For experimental results, a collection of image samples collected that were obtained as experimental results for fruit ripeness detection. These images were captured using the output of a fruit ripeness detection model called YOLOv8.

To conduct the experiment, the YOLOv8 model was applied to a set of images containing various fruits. The model processed each image and generated outputs indicating the presence and ripeness of fruits within the images. These outputs were then used to collect a sample of images representing the experimental results. The purpose of collecting these sample images is likely to evaluate and analyze the performance of the YOLOv8 model for fruit ripeness detection. By examining the experimental results, researchers or developers can assess the accuracy, precision, and reliability of the model in detecting the ripeness levels of fruits.

Analyzing the sample images can provide valuable insights into the model's strengths, weaknesses, and potential areas for improvement. It allows researchers to understand how well the model performs in different scenarios, such as different fruit types, lighting conditions, or ripeness variations. Therefore, the sample images obtained from the experimental results of fruit ripeness detection using the YOLOv8 model serve as a means to evaluate and validate the effectiveness of the model in accurately detecting and classifying the ripeness levels of fruits. Fig. 4 shows ripeness detection results.



Fig. 4. Fruit ripeness detection results.

### C. Performance Evaluation

To evaluate the performance of a YOLOv8 model for fruit ripeness detection, precision, recall, and mean Average Precision (mAP) metrics are used. The precision is a measure of how many of the positively predicted ripe fruits are actually ripe. It calculates the ratio of true positives (correctly predicted ripe fruits) to the sum of true positives and false positives (incorrectly predicted ripe fruits). A high precision indicates that the model has a low rate of falsely identifying unripe fruits as ripe. Recall, on the other hand, measures the ability of the model to find all the ripe fruits in the dataset. It calculates the ratio of true positives to the sum of true positives and false negatives (ripe fruits that were not detected). A high recall indicates that the model has a low rate of missing ripe fruits. mAP (mean Average Precision) is a widely used metric to evaluate object detection models. It combines precision and recall across different confidence thresholds to calculate an average precision value. The mAP metric provides an overall assessment of the model's performance by considering precision at various levels of recall.

In the given scenario, the precision of 98.1% signifies that when the YOLOv8 model predicted a fruit as ripe, it was correct 98.1% of the time. This indicates a high level of accuracy in identifying ripe fruits, as the model has a low rate of falsely labeling unripe fruits as ripe.

The recall value of 98.0% implies that the model successfully detected 98.0% of the ripe fruits present in the dataset. In other words, it only missed 2.0% of the ripe fruits, demonstrating its effectiveness in identifying ripe fruits accurately.

The mAP (mean Average Precision) score of 99.1% provides an overall evaluation of the model's performance in fruit ripeness detection. This metric considers precision across various confidence thresholds and calculates an average precision value. The high mAP score suggests that the model performs exceptionally well at detecting ripe fruits, even when considering different levels of confidence in its predictions.



Fig. 5. Results of precision, recall and mAP metrics.

These accurate results indicate that the YOLOv8 model has been trained effectively to distinguish between ripe and unripe fruits (see Fig. 5). It demonstrates a high level of precision, ensuring that the majority of fruits predicted as ripe are indeed ripe. Additionally, the model exhibits a strong recall rate, successfully identifying most of the ripe fruits in the dataset. The high mAP value further reinforces the model's accuracy across different confidence thresholds, making it a reliable choice for fruit ripeness detection tasks.

## V. CONCLUSION

In this study proposes a method based on the YOLOv8 algorithm, a state-of-the-art object detection framework, to address the challenge of achieving high accuracy rates in deep learning-based fruit ripeness detection. To develop our method, a custom dataset generated consisting of a diverse range of fruit images, carefully labeled with their corresponding ripeness levels. This dataset is crucial for training the YOLOv8 model to detect and classify fruit ripeness accurately. Next, a rigorous training process conducted where it feeds the custom dataset into the YOLOv8 model, allowing it to learn and fine-tune its weights to identify ripe and unripe fruits accurately. Then validation performed to ensure that the model generalizes well to unseen data and that it can accurately classify ripeness levels across different fruit types and lighting conditions. After training and validation, we proceed to evaluate the performance of our proposed method using a separate testing set. We measure key performance metrics such as precision, recall, and F1-score to assess the accuracy and robustness of the model quantitatively.

Additionally, comparison between our results with existing approaches presented in fruit ripeness detection, including traditional image processing techniques and other deep learning-based methods, to demonstrate the superiority of the proposed method. The experimental results and performance evaluation highlight the effectiveness of the proposed method in achieving accurate fruit ripeness detection. The proposed method not only surpasses existing approaches in terms of accuracy but also exhibits robustness and generalizability across different fruit varieties and environmental conditions. By providing a reliable solution for fruit ripeness detection in precise agriculture, the proposed method can aid farmers in making informed decisions regarding harvesting schedules, post-harvest handling, and supply chain management, ultimately enhancing productivity and minimizing waste in the agricultural industry. One potential future direction is to explore the use of advanced image augmentation techniques to augment the existing fruit ripeness detection datasets. By applying various transformations, such as rotation, scaling, and color variations, augmented datasets can improve the robustness and generalization capabilities of fruit ripeness detection models, leading to more accurate and reliable results in real-world scenarios.

## REFERENCES

- [1] I. Cisternas, I. Velásquez, A. Caro, and A. Rodríguez, "Systematic literature review of implementations of precision agriculture," *Computers and Electronics in Agriculture*, vol. 176, p. 105626, 2020.
- [2] R. Akhter and S. A. Sofi, "Precision agriculture using IoT data analytics and machine learning," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 8, pp. 5602-5618, 2022.
- [3] U. Shafi, R. Mumtaz, J. García-Nieto, S. A. Hassan, S. A. R. Zaidi, and N. Iqbal, "Precision agriculture techniques and practices: From considerations to applications," *Sensors*, vol. 19, no. 17, p. 3796, 2019.
- [4] R. Thakur, G. Suryawanshi, H. Patel, and J. Sangoi, "An innovative approach for fruit ripeness classification," in *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2020: IEEE, pp. 550-554.
- [5] S. Abasi, S. Minaei, B. Jamshidi, and D. Fathi, "Development of an optical smart portable instrument for fruit quality detection," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1-9, 2020.
- [6] B. N. Gharaghani, H. Maghsoudi, and M. Mohammadi, "Ripeness detection of orange fruit using experimental and finite element modal analysis," *Scientia Horticulturae*, vol. 261, p. 108958, 2020.
- [7] S. A. Ghazali, H. Selamat, Z. Omar, and R. Yusof, "Image analysis techniques for ripeness detection of palm oil fresh fruit bunches," *ELEKTRIKA-Journal of Electrical Engineering*, vol. 18, no. 3, pp. 57-62, 2019.
- [8] M. Rizzo, M. Marcuzzo, A. Zangari, A. Gasparetto, and A. Albarelli, "Fruit ripeness classification: A survey," *Artificial Intelligence in Agriculture*, 2023.
- [9] J. W. Lai, H. R. Ramli, L. I. Ismail, and W. Z. Wan Hasan, "Oil palm fresh fruit bunch ripeness detection methods: a systematic review," *Agriculture*, vol. 13, no. 1, p. 156, 2023.
- [10] B. Chen, M. Zhang, H. Chen, A. S. Mujumdar, and Z. Guo, "Progress in smart labels for rapid quality detection of fruit and vegetables: A review," *Postharvest Biology and Technology*, vol. 198, p. 112261, 2023.
- [11] A. P. Singh, P. Sahu, A. Chug, and D. Singh, "A Systematic Literature Review of Machine Learning Techniques Deployed in Agriculture: A Case Study of Banana Crop," *IEEE Access*, vol. 10, pp. 87333-87360, 2022.
- [12] S. Tulli and Yogesh, "Application of Machine Learning for Analysis of Fruit Defect: A Review," *Computational Intelligence: Select Proceedings of InCITe 2022*, pp. 527-537, 2023.
- [13] A. Septiariini, H. Hamdani, H. R. Hatta, and A. A. Kasim, "Image-based processing for ripeness classification of oil palm fruit," in *2019 5th International Conference on Science in Information Technology (ICSITech)*, 2019: IEEE, pp. 23-26.
- [14] Buitems. "Ripeness Detection FYP Project Zaid Dataset." <https://universe.roboflow.com/buitems-5ybt0/ripeness-detection-fyp-project-zaid>, 2023.
- [15] B. Xiao, M. Nguyen, and W. Q. Yan, "Fruit ripeness identification using transformers," *Applied Intelligence*, pp. 1-12, 2023.
- [16] D. Xu, H. Zhao, O. M. Lawal, X. Lu, R. Ren, and S. Zhang, "An Automatic Jujube Fruit Detection and Ripeness Inspection Method in the Natural Environment," *Agronomy*, vol. 13, no. 2, p. 451, 2023.
- [17] T.-Y. Lin et al., "Microsoft coco: Common objects in context," in *Computer Vision—ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, 2014: Springer, pp. 740-755.
- [18] A. Mehra. "Understanding YOLOv8 Architecture, Applications & Features." <https://www.labellerr.com/blog/understanding-yolov8-architecture-applications-features>, 2023.
- [19] Veeramsetty, Venkataramana, Bhavana Reddy Edudodla, and Surender Reddy Salkuti. "Zero-Crossing Point Detection of Sinusoidal Signal in Presence of Noise and Harmonics Using Deep Neural Networks." *Algorithms* 14.11 (2021): 329.
- [20] Veeramsetty, Venkataramana, D. Rakesh Chandra, and Surender Reddy Salkuti. "Short term active power load forecasting using machine learning with feature selection." *Next Generation Smart Grids: Modeling, Control and Optimization*. Singapore: Springer Nature Singapore, 2022. 103-124.

# A New Method for Intrusion Detection in Computer Networks using Computational Intelligence Algorithms

Yanrong HAO, Shaohui YAN\*

Department of Software Engineering, Hebei Software Institute  
Baoding 071000, China

**Abstract**—This paper introduces a novel and integrated approach to intrusion detection in computer networks that makes use of the benefits of both abuse detection and anomaly detection techniques. The proposed method combines anomaly detection and abuse detection technologies to enhance intrusion detection functionality. The intrusion detection system is implemented using a set of algorithms and models in the proposed approach. The frog jump algorithm has been utilized to choose the system's ideal input attributes. The decision tree is utilized in this system's abuse detection portion. Support vector machines or basic-radial neural network models have been utilized to find anomalies in this system. In the process of training neural networks, other techniques like particle swarm or genetic optimization are also utilized. The NSL-KDD dataset was used the experiment, and the findings were published. These findings demonstrate that, in comparison to using only anomaly or abuse detection, the proposed approach can increase the effectiveness of intrusion detection in the network. Additionally, a model that uses the frog leap algorithm for feature selection and classification and combines decision tree and support vector machine techniques with ten chosen input features has a detection rate of 98.2%. This is true despite the fact that the detection rates of the systems trained using comparable data in prior studies with 33 and 14 selected input features to the trainer have been 83.2% and 84.2%, respectively. Additionally, the algorithm execution performance increases up to 29 times faster than the aforementioned approaches when the intrusion detection rate is maintained at the level of other competing methods that were simulated in this work.

**Keywords**—Decision tree; network intrusion detection; particle swarm algorithm; basic-radial neural network; frog jump algorithm

## I. INTRODUCTION

Computers connected to the Internet are threatened by a variety of things, including unauthorized access to user systems and the execution of unpleasant behaviors [1]. Typically, network penetration is viewed as an attack [2]. Intrusion detection systems are already a commonplace component of the security architecture. The following is a list of the intrusion detection system's objectives: Preventing behavioral issues that attack or abuse the system, recognizing attacks and coping with them, documenting current attacks, quality control, and giving security managers meaningful penetration information are just a few of the objectives [3].

According to the "CSI/FBI Computer Security and Crimes Survey" report, intrusion detection system usage increased from 42% in 1999 to 62% in 2010 and will reach 62% by 2022 has retained. These numbers demonstrate the critical role that these systems play in security technologies [4].

Network intrusion detection systems monitor network activity to find assaults. Exploit detection and anomaly detection are the two primary techniques for intrusion detection [5]. Using patterns and signatures that signal assaults, you can find exploits and intrusions. Detects assaults but is unable to identify them [6]. Due to the use of less complex identification algorithms, the abuse detection technique has the advantage of extremely rapid identification [7]. Activities that depart from regular functioning are identified as infiltration in the anomaly detection approach by the construction of normal usage profiles. Because of this, the exploit detection approach is unable to identify unexpected intrusions that the anomaly detection system can [8]. The high prevalence of false alarms in the anomaly detection approach is one of its shortcomings [9].

Intrusion detection systems that integrate both strategies have been developed to address the issues with these two approaches [10]. These systems do it in three different ways. The following approaches are used: a. abnormality is first identified, followed by abuse; b. parallel approach; and c. abuse is first identified, then abnormality [11].

The detection rate of all sorts of assaults (known and unknown) increases because under the parallel approach, incoming traffic is evaluated independently by each method (identification of abuse and identification of anomalies) [12]. The anomaly detection approach still has a high risk of false positive notifications, but if the detection model qualifies the communication as an attack, this method likewise treats the incoming communication as an assault [13, 42]. The computational cost of detection is another concern, as each communication must be examined using both anomaly detection and abuse detection models, which raises this cost [14].

In the combined strategy employed in this article, we have attempted to produce an ideal plan by combining already-existing algorithms and models [15]. In this regard, it has been managed to minimize the number of input features to the system from 41 to 10 thanks to the deployment of the optimization technique based on frog jump. Comparing this



strategy to several studies in this subject, the combined system is also one of the drawbacks. In this manner, the abuse detection system receives the inbound traffic first (Fig. 1). The exploit detection phase's outputs that do not fit the intrusion patterns are fed into the anomaly detection system as input in order to find unidentified intrusions [16, 43]. The effectiveness of anomaly detection declines as the volume of attacks rises. Hence exploit detection has been used first to address this issue. Known attacks can be found using exploit detection. The quantity of attacks required to find anomalies is drastically decreased by eliminating known attacks. Another benefit is that the suggested hybrid system can identify known intrusions in real-time due to the high speed of exploit detection techniques such as decision tree algorithms [17].

In this situation, the plan's anomaly detection component uses well-established, successful models (such as artificial neural networks) that can recognize novel attacks [6]. Instead of employing conventional methods for neural network training, computational intelligence algorithms have been adopted since they have proven to be more effective [18].

The main motivation of this research is to improve the effectiveness of intrusion detection systems (IDS). By combining abuse detection and anomaly detection methods, this approach aims to increase the system's ability to detect and mitigate various types of network intrusions. Traditional IDSs may struggle to effectively identify both known and unknown threats. Another motivation is to optimize the use of computing resources and reduce processing time. The choice of algorithms such as frog jumping algorithm, decision trees and support vector machines (SVM) shows the desire to achieve efficient and accurate intrusion detection without computational overhead. In summary, the proposed approach in this research

is motivated by the need for more effective intrusion detection systems; more efficient and adaptable in the face of evolving cyber threats. Potential benefits include improved detection rates, resource optimization, reduced false positives, and increased consistency, all of which contribute to a stronger and more comprehensive network security solution. In this regard, the suggested method's usage of radial basis neural network (RBF) in the anomaly detection phase has also led to a notable increase in the algorithm's execution speed when compared to other hybrid techniques. A comparison of the proposed hybrid system's performance with other similar schemes that have been tried on the same field with related events reveals that the suggested scheme has attained the desired detection rate (see Table VII in Section V). The main contributions of this research and their possible consequences are as follows:

- Hybrid intrusion detection system: This research introduces a new hybrid intrusion detection system that combines abuse detection and anomaly detection techniques. This hybrid approach can lead to more effective and robust intrusion detection because of its strengths. Both methods are used. The result is improved network security and a greater likelihood of detecting a wide range of intrusions.
- Feature selection and data preprocessing: This research emphasizes the importance of feature selection and data preprocessing techniques to improve the quality of input data. Choosing the right feature and normalizing the data can lead to a more reliable and accurate intrusion detection system. The consequences are higher accuracy in detecting intrusions and reduction of noise in the data.

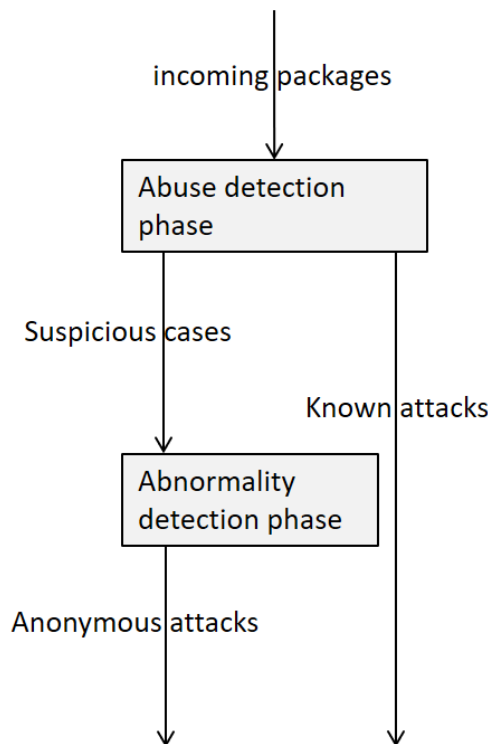


Fig. 1. Hybrid intrusion detection system with abuse detection priority.

- **Efficiency and scalability:** This research deals with the efficiency and scalability of the intrusion detection system, especially by comparing SVM and radial basis neural networks. Understanding the trade-offs between different algorithms and their impact on performance and scalability can help deploy intrusion detection systems in different network environments. The result is consistent and resource-efficient intrusion detection.

The remainder of the essay is structured as follows. Studying earlier works is the subject of Section II. Section III offers suggestions for spotting abuse and abnormalities. Section IV discusses evaluation and simulation results, and Section V concludes with recommendations for future research.

## II. BACKGROUND RESEARCH

This section will first cover some earlier research on hybrid systems before introducing the fundamental models and techniques employed in this study.

### A. Related Work

Some of the studies on hybrid systems have been covered in this subsection. The three-layer architecture of the intrusion detection system, which was created and developed by [19], is an illustration of a hybrid system. Based on the KDD Cup'99 standard data, the system featured a blocklist, an allowlist, and a multiclass support vector machine bundle. Blocklists are used to filter out known threats from network traffic, whereas allowlists are used to identify legitimate traffic.

A method for audit data mining and analysis (ADAM) was presented by [20] in which abuse detection is used after anomaly detection. In order to identify attacks, ADAM combined an association mining rule with a classification mechanism. The suspect traffic is first identified by the association mining rule-based anomaly detection model, which then sends it to the abuse detection model. After then, the suspicious communications are classified by the abuse detection model as "normal," "known attacks," and "unknown attacks" (false alert of the anomaly detection model). Its usage is uncommon. Communications that cannot be categorized as typical patterns or known attacks under the ADAM technique are categorized as unknown attacks. If anomaly detection is used first, then abuse detection, the anomaly detection model should have a high detection rate, and the abuse detection model should eliminate the false alarms generated by the anomaly detection model by differentiating between known and unknown attacks. The majority of abuse detection systems, however, are ineffective at lowering false alerts. An anomaly detection model, an abuse detection model, and a decision support system were all incorporated in the [21] proposal for an intelligent hybrid intrusion detection system. They used a decision tree to model the abuse detection model and a self-organizing map (SOM) neural network to model anomaly detection. Following independent training of each model, the decision support system pooled the categorization outcomes of the two models.

A hybrid intrusion detection system was designed in [22] using the abuse detection approach first, then the anomaly detection method. The exploit detection model is quicker than

the anomaly detection model and can identify known attacks with a low probability of false positives. In order to identify known attacks, the Al-Teda detection and exploitation model was employed. Only non-deterministic connections were then detected using the anomaly detection model. A technique for detecting anomalies identifies outliers that deviate from typical data patterns and classifies them as well-known assaults. The anomaly detection and abuse detection models, however, are trained independently, just like the parallel hybrid technique, which causes a high risk of false positive notifications in the outcomes.

With a hybrid methodology, researchers in [23] initially used the C4.5 decision tree algorithm to evaluate the traffic during the abuse detection phase. After the exploit detection phase, in the anomaly detection phase, distinct support vector machines (SVM) were employed to discover unexpected incursions for each subset of data classified as normal by the intrusion detection model. Each subset will be more effective at generating typical profiles and finding anomalies because it has more concentrated data.

Computational intelligence techniques have been utilized for feature selection and decision trees for classification in numerous studies in the area of computer network security. For instance, researchers in [24] utilized the decision tree with the C4.5 training method to identify garbage items and the binary version of the particle swarm optimization technique (BPSO) for feature selection. The ideal collection of features was likewise chosen by the source [25] using CFA-based optimization, and the chosen features were assessed using a decision tree-based classifier. The estimation and selection of acceptable and chosen features for data classification during the tree training process is one of the decision tree's beneficial properties. On the other hand, adding a feature selection phase before training the decision tree has been shown in experiments to significantly improve the classification accuracy of the decision tree [26]. As a result, the frog jump method incorporated a feature selection step before training the decision tree in the article's suggested solution. The techniques and models used in this article are briefly introduced in the sections that follow so that the next sections can explore how to combine them and express the simulation results.

### B. Frog's Leap Algorithm

The combined optimization method based on frog jumping (SFLO) is one of many evolutionary algorithms that have been created in recent decades to decrease processing time and increase the quality of results [24]. This program uses a technique called imitative discovery [27] and aims to use a heuristic search to identify the overall best answer. In order to discover effective general solutions, this technique has been tested on a number of combinatorial problems. The population of potential solutions for the frog leaping algorithm is defined as a set of subsets of frogs (solutions). Distinctive subgroups are viewed as distinct frog species, each of which conducts a distinctive local search. Each frog in the subgroups has ideas that are shaped by those of other frogs and changed by the process of imitational evolution. Ideas spread through the subsets through the process of interweaving and interweaving after going through the evolutionary phases of imitation. Until

the convergence rule is satisfied, local search and nesting operations are carried out [28].

### C. Decision Tree

One of the most well-known techniques for creating a classification model is the decision tree. The result information of decision tree-based classification algorithms is displayed as a tree of several feature value states. Always dividing records based on a candidate feature that maximizes a particular criterion is a greedy approach to decision tree construction. The most deserving feature will be the one that improves the tree the most in accordance with this criterion. Depending on how the features of the data set are chosen to be included in the decision tree and when to cease growing the tree, there are various types of decision trees. Better than other failures are one where the distribution of bundles in the resulting nodes is homogeneous. The node is homogeneous if all of its records are suspended in the same category. Since, in this situation, that node turns into a leaf. In actuality, the node with the least number of impurities is the homogeneous node. They are placed in the interest relationship, and the amount of interest resulting from each failure is calculated after the amount of impurity arising from each failure has been determined. The failure-related impurity is removed from the parent node's impurity in the gain relation. Any failure that generates a greater profit is preferable, and that failure will ultimately be chosen [26]. The C4.5 tree, which is the decision tree employed in this article, is utilized to determine its impurity using the entropy approach based on Eq. (1):

$$\text{Entropy}(t) = - \sum_i p(j|t) \log p(j|t) \quad (1)$$

In relation (1),  $p(j|t)$  denotes the proportion of records belonging to the  $j$ th category to all other records in node  $t$ . The

fracture gain is calculated after the entropy for each node has been determined, followed by the entropy for the entire fracture. A failure is better if it has a greater interest rate. The gain of a failure is calculated using Eq. (2):

$$\text{GAIN}_{\text{split}} = \text{Entropy}(p) - \sum_{i=1}^k \frac{n_i}{n} \text{Entropy}(i) \quad (2)$$

Regarding this,  $n$  represents the overall number of records in the parent node,  $n_i$  represents the number of records in the  $i_{\text{th}}$  child,  $\text{entropy}(p)$  represents the entropy of the parent node, and  $\text{entropy}(i)$  also represents the entropy of the  $i_{\text{th}}$  node [29]. The test data set can then be used to test the classification model that was created using the decision tree. The goal of using the model is to predict, using the model, the category attribute value for the test record.

### D. Basic-Radial Neural Network (RBF)

The artificial neural network has the benefit of generalization, which sets it apart from other classifiers. With a small amount of training data, radial basis function networks are frequently used to estimate multidimensional functions nonparametrically. The RBF network is highly helpful since it trains quickly and thoroughly. With enough neurons in the hidden layer, it can estimate any continuous function with any level of accuracy. Fig. 2 depicts the three-layer network that makes up RBF's main design.

The basic-radial functions are shared by the neurons in the intermediate (hidden) layer [30]. According to Eq. (3), the third layer approximates the middle layer neurons' output by summing their weighted output:

$$F(x) = \sum_{i=1}^p w_j \phi(||x - u_j||) \quad (3)$$

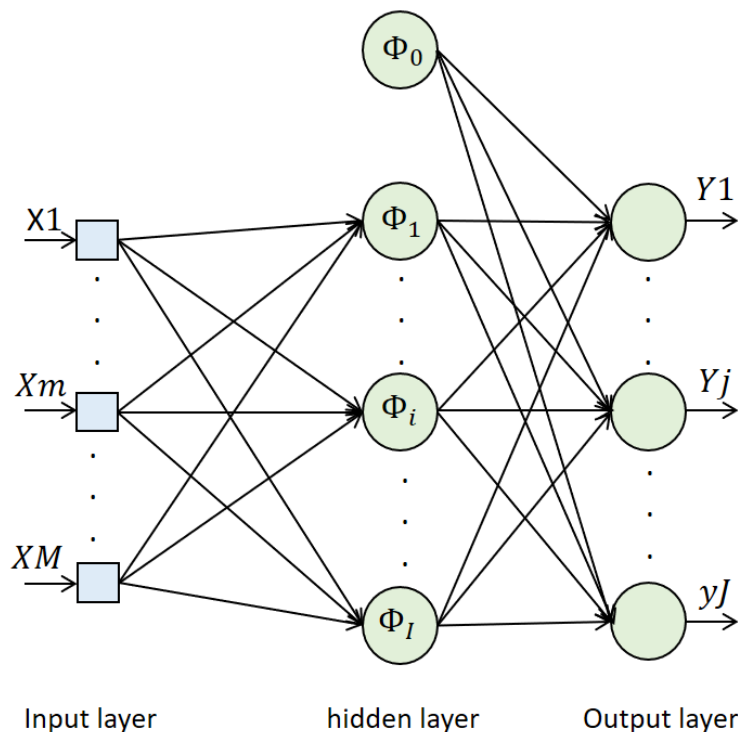


Fig. 2. The architecture of an RBF network [30].

If RBF is used to approximate the function, its output will be helpful. However, a hard limiter can be applied to the output neurons to ensure output values of 0 or 1 for pattern classification purposes. The base-radial function  $p$  with centers  $u_j$  is utilized for the approximation of the F function based on Eq. (3). Sign  $\|\cdot\|$  the exponent, which is typically set to represent the Euclidean distance, is the distance function of  $R_n$  space. The most well-known basic-radial function, which has been proposed, is the Gaussian function in RBF networks. The Gaussian function, an exponential function from the category of functions with the best approximation qualities, was chosen as the response function of neurons in RBF networks. This ensures that there is a set of weights that more accurately approximates the relationship between the input and target vectors than any other set. The sigmoid function utilized in the creation of error backpropagation networks makes use of this assurance that it has no existence.

### E. Particle Swarm Optimization Algorithm

The social behavior of a group of birds is described by the population-based stochastic optimization algorithm known as the PSO algorithm. In space, a flock of birds looks for food randomly. Following the bird that is closest to the food can be one of the greatest tactics. The PSO algorithm's core concept is this tactic [31]. The search space for the PSO algorithm is analogous to the search space for the bird movement pattern. In the PSO algorithm, each solution, also known as a particle, is analogous to a bird, and there are exactly as many particles (solutions) as there are birds. Each particle has a merit value, which is determined by a merit function. The more merit a particle has, the closer it is to the target in the search space, which in the bird movement model is food. Additionally, every particle has a displacement that controls its direction of motion and is used to predict its next location. Until it ultimately reaches the ideal solution, each particle moves forward in the search space by adhering to the best particles in the current state [32]. The particle velocity vector ( $V_i$ ) and particle ( $X_i$ ) in the  $(t+1)$ th iteration are computed from relations (4) and (5) in each step of the particle swarm algorithm iteration:

$$\begin{aligned} V_i(t+1) &= w \times V_i(t) + \\ &C_1 \times \text{rand}_1 \times [p\text{best}_i(t) - X_i(t)] + \\ &C_2 \times \text{rand}_2 \times [g\text{best}_i(t) - X_i(t)] \quad (4) \\ X_i(t+1) &= X_i(t) + V_i(t+1) \quad (5) \end{aligned}$$

### F. Hereditary Algorithm

Hand first proposed the genetic algorithm in 1965. From among the chromosomes in a population, the selection operator chooses the number of chromosomes for reproduction. Fitter chromosomes are more likely to be chosen for reproduction. Chromosome segments are randomly switched between during crossing over. Because of this, the children don't exactly resemble one of their parents but instead exhibit traits from both. A single-point, two-point, or even intersection could exist [33].

All chromosomal points have an identical chance of merging during the uniform crossover. In this case, any valley can be used to pick the child's chromosome genes. It is possible to perform uniform intersection using relations (6) and (7):

$$y_{1i} = a_i x_{1i} + (1 - a_i) x_{2i} \quad (6)$$

$$y_{2i} = a_i x_{2i} + (1 - a_i) x_{1i} \quad (7)$$

In these relationships,  $x_1$  and  $x_2$  are the parents, while  $y_1$  and  $y_2$  are the first and second children, respectively. Additionally, the values of  $a_i$  in continuous issues are in the range  $[0,1]$ , while those in binary problems are equal to zero or one. The letter  $i$  stands for the input dimensions (solution space dimensions).

The chromosomes are subjected to the mutation operator after crossing over. This operator chooses a gene at random from a chromosome and modifies the information within that gene. If the gene is a binary number, it is inverted; if it is a member of a set, another value or component of that set is substituted for the gene. The created chromosomes are known as the new generation and are sent to the following round of algorithm execution after the mutation operation is finished.

## III. SUGGESTED METHOD

Fig. 1 shows the suggested method's operating principles. The specifics of spotting abuse and abnormalities will be covered in this section. Prior to applying the data to the model in the proposed combined method, the data was first preprocessed using the leapfrog computational intelligence algorithm to extract its key features, which improved the system's overall performance and, in particular, the effectiveness of the decision tree. The regular training data is then separated into subgroups during the misuse detection stage whose connection patterns have less variation than the entire normal data. Then, in the anomaly detection stage, a different anomaly detection model is applied for each subset. As a result, each subset's efficiency will be higher in producing more normal profiles and rejecting the result in anomaly detection since each subset has more concentrated data [9, 34]. The quality of the input data to the anomaly detection stage is not good enough and has a significant impact on the accuracy of the SVM network if the decision tree is unable to appropriately partition the data into subsets while retraining the model in the operational environment. This lowers the effectiveness of the SVM network; however, the RBF neural network is not constrained by this issue. Since model training is a linear process and neural network training is slower, the effectiveness of the models is unaffected [34].

The RBF neural network was employed in the anomaly detection phase, and the effectiveness of SVM and RBF was compared in the anomaly detection phase, in accordance with the foregoing and to prevent the reduction of the accuracy of the SVM network, in cases where the output of the decision tree is not effective. It is important to note that switching from SVM training to RBF training and performing feature selection prior to the abuse detection stage in order to condense the problem's dimensions were the adjustments that significantly increased the effectiveness of the newly proposed method. The simulations were conducted using Matlab software version 2022a as well. The steps for intrusion detection in the suggested approach are as follows:

- Data preprocessing includes the following steps:

a) Data homogenization (training and test data) by replacing the letter characters of the data set with numerical values

b) Data normalization (training and test data): In this step, the values of the continuous features are normalized to the interval [-1,1] according to Eq. (8).

$$X = 2 \times \frac{x - \min(x)}{\max(x) - \min(x)} - 1 \quad (8)$$

c) Reducing the dimensions of the problem by feature selection by the leapfrog computational intelligence algorithm

It should be noted that normalization (or rescaling of features) removes the imbalance between the data [36-34]. In the dataset used in this paper (NSL-KDD), some features have large numerical values that can dominate other features. For example, the *dst\_bytes* attribute can have values from zero to about  $1.3 \times 10^9$ , while the *same\_srv\_rate* attribute has values between zero and one [35].

- Applying various methods in the combined model of penetration detection:

(First the abuse detection phase and then the abnormality detection phase) including the C4.5 decision tree in order to detect abuse and then apply separate RBFs in the abnormality detection phase for the categories that are detected as normal. It is reminded that the training parameters in the RBF neural network are determined with the help of the genetic algorithm or the particle swarm algorithm. The block diagram presented in Fig. 3 shows the working steps in the proposed intrusion detection method in more detail.

### A. The Stage of Identifying Abuse

In the data mining process, the data are ready to be applied to the model learning stage after feature selection and data pretreatment. The C4.5 decision tree is utilized for this. The data exploration strategy that is chosen will determine the sequence guiding the preprocessed data during the learning stage, and the created model will then be sent to the evaluation stage (i.e., evaluation and interpretation of the model) for evaluation. It was time to cut the decision tree after training and preserving it, leaving only 12 leaves with conventional labels. Too small sets make this operation excessively slow because it takes time to split training data into distinct subsets based on various rules. The tree was pruned for these reasons because, on the other hand, the restriction and lack of overdispersion of neural network inputs during the training phase will reduce the generalization capacity of the network.

### B. Anomaly Identification Stage

At this point, distinct RBFs were utilized for each of the remaining leaves with the conventional label, and the data input for each of them was identical to the information classified by the decision-making process. The neural network's leaves had come. The conditions of their development were determined for all the leaves with the normal label in order to obtain the data in each leaf. Based on these conditions, the training data set was then segmented into several input subsets, and for each of them, different neural networks were employed. The usage of a more homogeneous data set boosts the network's accuracy, and the anomaly detection system is better prepared to deal with anomalies that do not follow the typical pattern since it is more accustomed to normal patterns with lower dispersion [36].

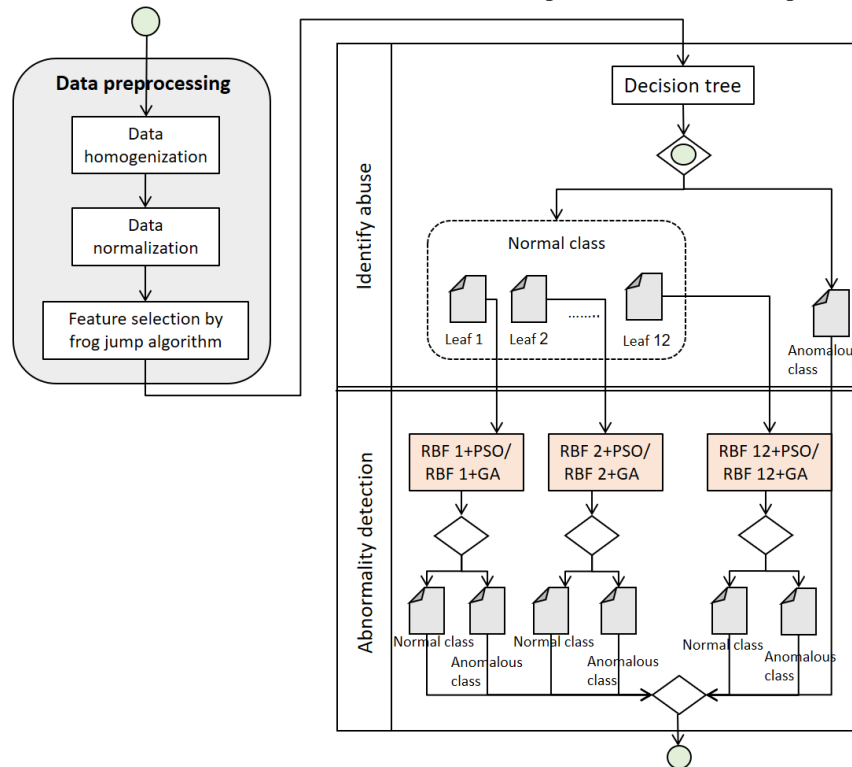


Fig. 3. Process block diagram of the proposed intrusion detection method.

#### IV. EVALUATION RESULTS OF THE PROPOSED SOLUTION

##### A. NSL-KDD Dataset

The NSL-KDD dataset has been utilized to evaluate and contrast the performance of the suggested method [37]. The NSL-KDD dataset is an edited version of the KDD'99 dataset that was made available as a result of KDD'99's issues with the presence of numerous duplicate samples in the training and test data, as well as the resolution of these issues. The 41 characteristics in this dataset span a wide range of numerical values and are of continuous, discrete, and symbolic types. As previously stated, data preprocessing is required to correct the imbalance in the entire dataset and eliminate the impact of scale differences in such a database [38]. In other words, normalization in such data on all data (in the following, some of them as training data and others as training data) tests are selected) [39] in order to prevent features with high numerical values from overpowering other features.

##### B. Overall Evaluation Results

The frog jump algorithm was used to intelligently choose ten attributes from among them in order to decrease the problem's dimension. Feature selection (variable reduction) aids in data comprehension, lower processing demands, lessens the impact of the dimensionality problem and enhances prediction performance. The goal of feature selection is to choose a subset of input variables that can accurately characterize the input data while minimizing the impact of extraneous factors and producing accurate prediction results [40].

In this step, the trained decision tree was first given all the input data from the abuse detection phase (signature detection) in order to evaluate the model. Assuming they are unidentified

attacks whose signatures or data packet characteristics could not be recognized by the exploit detection model; the identification component will once more detect the data that the decision tree classified as regular packets. In this phase, anomalies (trained RBF) were examined to determine their class. The way the anomaly detection phase operates is that the inputs are first reviewed again in accordance with the requirements of the decision tree's leaves to choose which of the RBFs should be provided as input. The selective RBF then determines the data packet's final class.

The time needed to run the model is significantly reduced if RBF is used in place of SVM, according to the results, so that the time needed for the test is decreased by an average of 26 times (if PSO is employed), a 24 times improvement, and if we employ GA, we will see an improvement in execution time of more than 28 times); however, the model error had a modest decline.

It should be noted that the ideal number of kernels for each RBF was found independently during the RBF training procedure. If PSO and inheritance algorithms are applied, the number of optimal cores for every RBF may be deduced from Fig. 4 and 5, respectively. According to the number of various kernels, each line in the graphs depicts the MSE error trend for one of the RBFs, and the number of optimal kernels for each RBF is the same as the number of kernels that minimize the MSE error.

As previously indicated, PSO and genetic algorithms (GA) were used to train the RBF neural network individually, and the outcomes of both were compared. The results were nearly identical, and no discernible difference was found regarding the effectiveness of applying PSO or GA in RBF training.

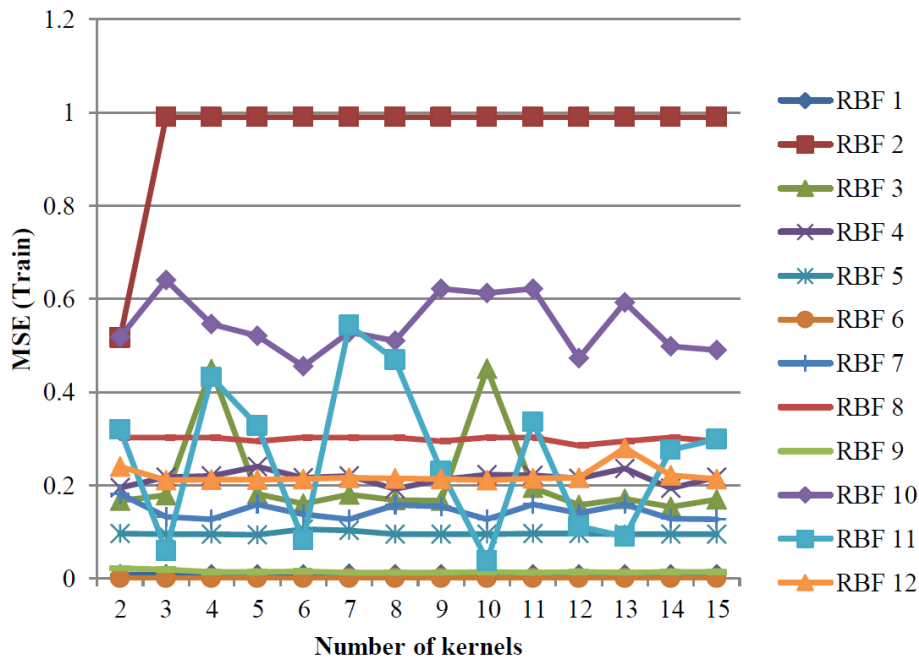


Fig. 4. Training error of all RBFs using PSO algorithm for different numbers of kernels.

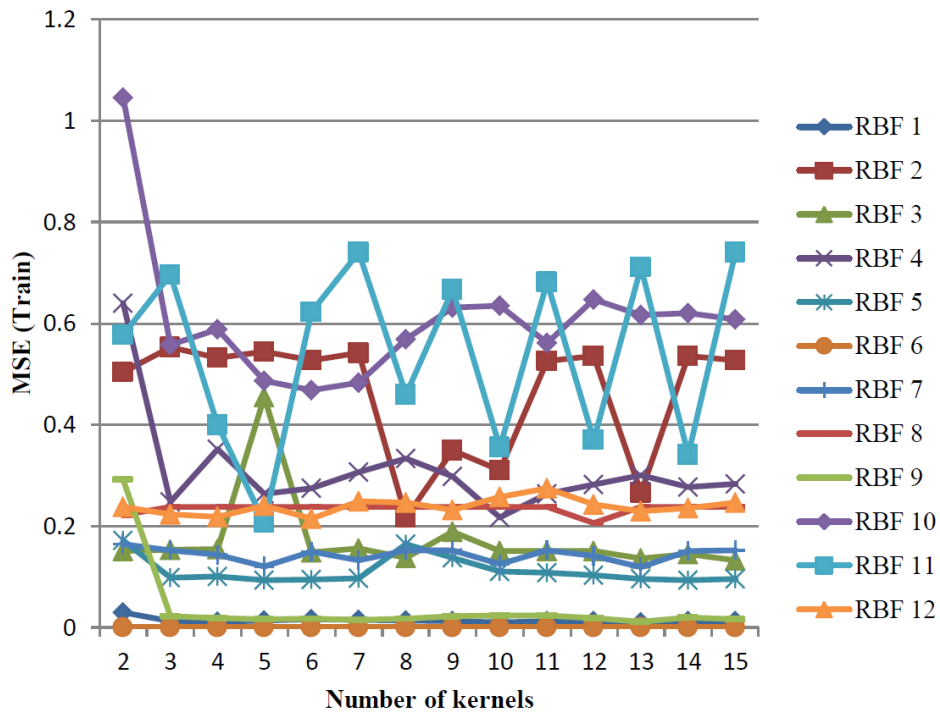


Fig. 5. Training error of all RBFs using GA for different numbers of kernels.

In accordance with relations (9) to (12), the following metrics were employed to assess the models: mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), and set of square error (SSE).

$$MSE = \sum_{i=1}^n \frac{(t_i - y_i)^2}{n} \quad (9)$$

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(t_i - y_i)^2}{n}} \quad (10)$$

$$MAE = \sum_{i=1}^n \frac{|t_i - y_i|}{n} \quad (11)$$

$$SSE = \sum_{i=1}^n (t_i - y_i)^2 \quad (12)$$

In the relationships mentioned above, n denotes the number of samples, t<sub>i</sub> is the goal output value, and u<sub>i</sub> denotes the output value of the class that the model predicted [41]. An Intel Core i7 machine with a clock speed of 2.1 GHz and 6 GB of random-access memory was used for the models' training and testing phases. Sixty-two thousand eight hundred sixty-four data points were used to train the models, while 85347 data points were used to test the models. The outcomes of using the models are shown in the paragraphs that follow.

### C. Decision Tree Training Results

Table I shows the results of errors in detecting abuse. The time required to train the tree was equal to 0.18 seconds.

### D. The Results of the Combined Model of Decision Tree and RBF with Training by PSO Algorithm

In Table II, the results of using the combined intrusion detection model are presented in the condition that the RBF model is trained with the PSO algorithm. The errors presented in Table II were obtained for this setting of values for the PSO algorithm parameters: number of population members: 60, personal learning coefficient (C1): 1.4962, collective learning coefficient (C2): 1.4962, W coefficient: 1 and Reduction factor W: 0.9.

### E. The Results of the Combined Model of Decision Tree and RBF with Training by GA

Table III also shows the results of using the combined intrusion detection model in the conditions where the RBF model is trained with the genetic algorithm. The errors presented in Table III were obtained for this setting of values for the parameters of the genetic algorithm: number of population members: 20, crossover probability: 0.9, mutation probability: 0.3, and mutation rate: 0.6.

TABLE I. ERROR-VALUES OF TRAINING AND TESTING IN THE PHASE OF DETECTING ABUSE BY DECISION TREE

Phase	MAE	MSE	SSE
Education	0.00200	0.00401	252
Test	0.1017	0.2034	17360

TABLE II. TRAINING ERROR VALUES OF INDIVIDUAL RBFs – TRAINING BY PSO ALGORITHM

Model	MAE	RMSE	MSE	SSE
RBF 1	0.00376	0.08677	0.00753	196
RBF 2	0.25846	0.71897	0.51692	12636
RBF 3	0.07680	0.39190	0.15359	828
RBF 4	0.09631	0.43888	0.19262	4228
RBF 5	0.04696	0.30646	0.09392	760
RBF 6	0.00011	0.01499	0.00022	4
RBF 7	0.06345	0.35622	0.12690	2744
RBF 8	0.14286	0.53452	0.28571	136
RBF 9	0.00591	0.10872	0.01182	680
RBF 10	0.22767	0.67480	0.45535	928
RBF 11	0.01862	0.19299	0.03724	20
RBF 12	0.10501	0.45827	0.21001	76

TABLE III. TRAINING ERROR VALUES OF DISCRETE RBFs-TRAINING BY GA

Model	MAE	RMSE	MSE	SSE
RBF 1	0.00536	0.10350	0.01071	140
RBF 2	0.10873	0.46632	0.21745	2664
RBF 3	0.06645	0.36454	0.13289	360
RBF 4	0.10859	0.46604	0.21719	2388
RBF 5	0.04647	0.30485	0.09293	376
RBF 6	0.00023	0.02123	0.00045	4
RBF 7	0.34446	0.34446	0.11865	1276
RBF 8	0.10317	0.45426	0.20635	52
RBF 9	0.00598	0.10938	0.01196	344
RBF 10	0.23392	0.68399	0.46784	480
RBF 11	0.10370	0.45542	0.20741	56
RBF 12	0.10724	0.46312	0.21448	400

Additionally, Table IV provides the amount of time needed to train the simulated models. Because PSO and GA have different populations with different numbers of population members, RBF training and PSO and GA differ in innovative ways. Each model with the least population members that leads to the lowest error rate has been taught since increasing the number of population members in population-based optimization algorithms increases training time because more iterations are required for more population members. As can be seen, using SVM rather than RBF cuts down on training time during the anomaly detection stage. This is because RBF training involves using the aforementioned smart optimization algorithms to find the ideal network weights, and finding these ideal values requires a significant amount of repetition. However, because the training process takes place offline, it is

acceptable to extend it in order to enhance the performance of the model's online execution.

#### F. The Results of Testing the Models with Test Data

The experimental results demonstrate that the model performs poorly when used in the anomaly detection phase with a significant amount of input data, which eventually results in the model's inefficiency in online applications (Table V). As can be shown, using a decision tree and an RBF neural network together (with training via the PSO algorithm) is the optimum option in terms of the model's online execution time. The aforementioned alternative offers competitive values and comes quite near to the combined decision tree and RBF neural network model (with GA training) in terms of the number of various error criteria (Table VI).

TABLE IV. THE TRAINING TIME OF THE MODELS

proposed model	Anomaly detection model training time (sec)	Time required to provide data for anomaly detection model (sec)	Training time of the misuse detection model (sec)	total time (sec)
C4.5+SVM	29.48314	17.79945	0.18482	47.4641
C4.5+RBF-PSO	78.76758	17.79945	0.18482	96.75185
C4.5+RBF-GA	59.21933	17.79945	0.18482	77.20360



TABLE V. TESTING TIME OF MODELS WITH TEST DATA

proposed model	total time (sec)
C4.5+SVM	77.216326
C4.5+RBF-PSO	2.701057
C4.5+RBF-GA	3.229697

TABLE VI. THE TESTING ERROR OF PROPOSED MODELS WITH TEST DATA

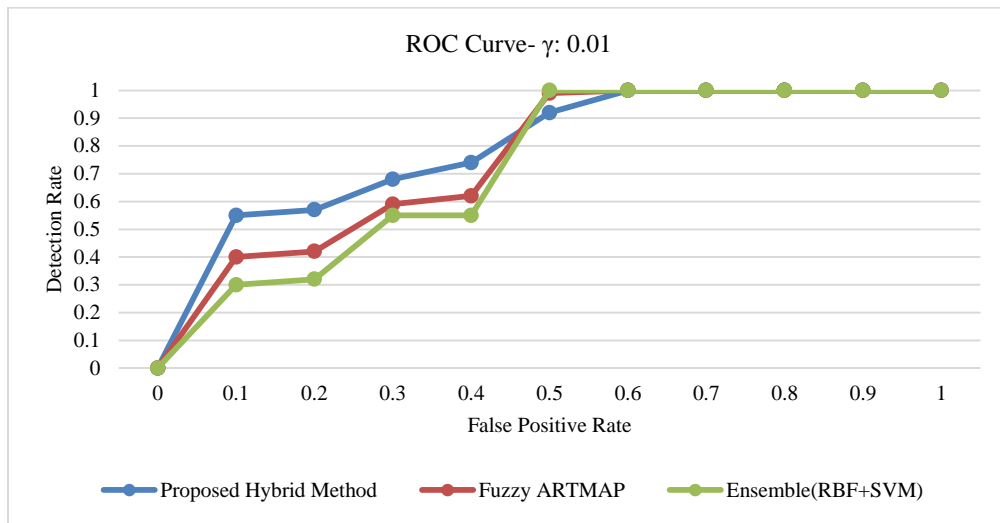
proposed model	MAE	RMSE	MSE	SSE
C4.5+SVM	0.17470	0.59110	0.3494	29820
C4.5+RBF-PSO	0.16685	0.57766	0.3337	28480
C4.5+RBF-GA	0.16645	0.57697	0.3329	28412

TABLE VII. COMPARISON OF THE DETECTION RATE OF THE PROPOSED SYSTEM WITH SIMILAR SYSTEMS

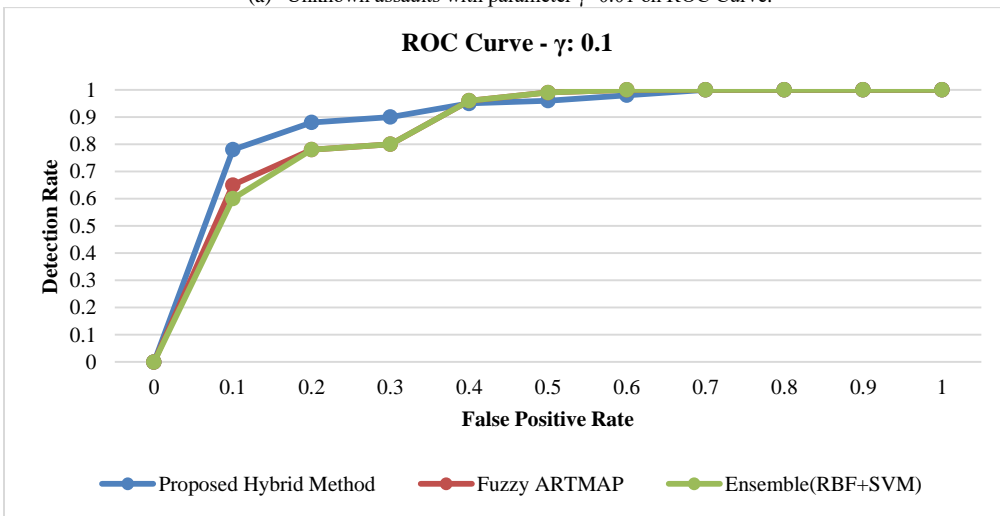
Abbreviated name of the Amikhtar model	Feature selection algorithm	Classification tool	Data Names	Number of selected features	Detection rate (%)
FGBARM <sup>1</sup> + GSA <sup>2</sup> – Fuzzy ARTMAP[28]	FGBARM	GSA-Fuzzy ART MAP	KDD'99	26	93.74
FGBARM+PSO-Fuzzy ARTMAP[31]	FGBARM	PSO-Fuzzy ART MAP	KDD'99	29	97.25
FPGBARM+PSO-Fuzzy ARTMAP[32]	FGBARM	GA-Fuzzy ART MAP	KDD'99	29	9790
DBN <sup>3</sup> + SVM[33]	DBN	SVM	NSL-KDD	14	83.14
RST <sup>4</sup> + SVM[34]	RST	SVM	KDD'99	29	89.36
Info-Gain+J48[27]	Info-Gain	J48 decision tree	NSL-KDD	33	81.94
Info-Gain+Natve Bayes[36]	Info-Gain	Natve Bayes	NSL-KDD	33	75.79
Info – Gain + MLP <sup>3</sup> [18]	Info-Gain	MLP	NSL-KDD	33	73.55
Info-Gain+SVM[41]	Info-Gain	SVM	NSL-KDD	33	71.02
Info – Gain + CART <sup>6</sup> [42]	Info-Gain	CART	NSL-KDD	33	82.32
Ensemble+Bayesian[43]	Applying nine basic features, 13 content features, and 19 traffic features to three separate categories	Fuzzy K-NN, MLP, and Naive Bayes classifiers	KDD'99	A total of 41 features are used in three categories	93.35
Ensemble (RBF+SVM)[21]	Best First Search	RBF+SVM	NSL-KDD	9	85.19
SFLO+C4.5-SVM (recommended model)	SFLO	C4.5-SVM	NSL-KDD	10	97.4
SFLO+C4.5-PSO-RBF (recommended model)	SFLO	C4.5-PSO-RBF	NSL-KDD	10	96.9
SFLO+C4.5-GA-RBF (recommended model)	SFLO	C4.5-GA-RBF	NSL-KDD	10	96.9

With a false positive rate of 1.3%, DT has a detection rate of 99.2% for known threats and 31.06 percent for unknown attacks. DT turns out to be unsuitable for detecting new assaults but has strong detection performance for known attacks. The proposed method's detection performance was carefully examined in terms of anomaly detection since its goal is to enhance the detection performance of the anomaly detection model. Fig. 6 compares the suggested method's receiver operating characteristic (ROC) curves for unknown attacks with those for increases in parameter  $\gamma$  from 0.01 to 1. The proposed method has a greater detection rate than traditional ones. The rate of false positives in the decision boundaries of the anomaly detection model in the proposed method can depict normal behavior better than existing methods since each class 1 SVM model can concentrate on its corresponding decomposed region.

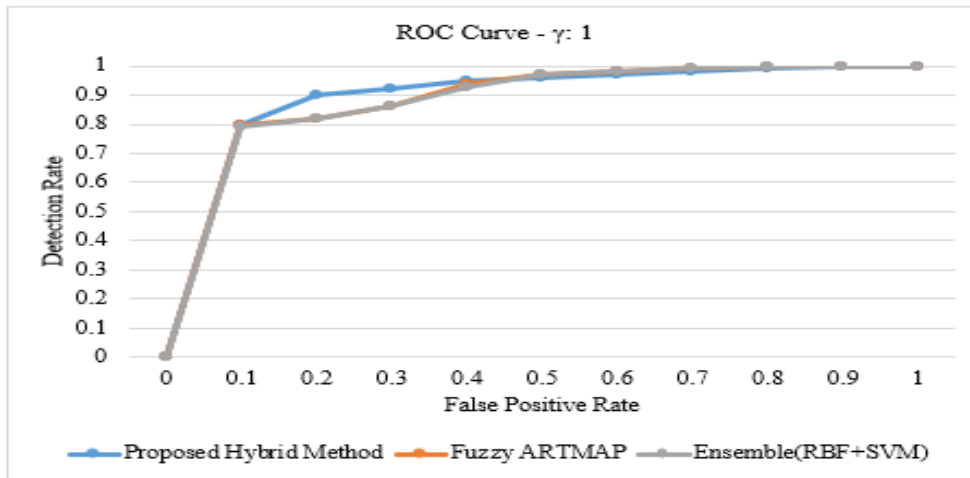
The proposed method's detection rate is roughly 11% greater than that of traditional approaches when the false positive rate is below 11%, which is ideal. It was found that the conventional method's detection rate improves when the false positive rate is about 51%. This outcome is believed to be the consequence of the suggested technique, which calls for building a class 1 SVM model for each deconstructed region. Instead of concentrating on each deconstructed zone when the false positive is very significant (like 51%), it is preferable to concentrate on the highly concentrated regions. It can be deduced that the detection performance of the suggested method is superior to traditional methods for unknown assaults because an IDS operator should have a very low false positive rate.



(a) Unknown assaults with parameter  $\gamma=0.01$  on ROC Curve.



(b) Unknown assaults with parameter  $\gamma=0.1$  on ROC Curve



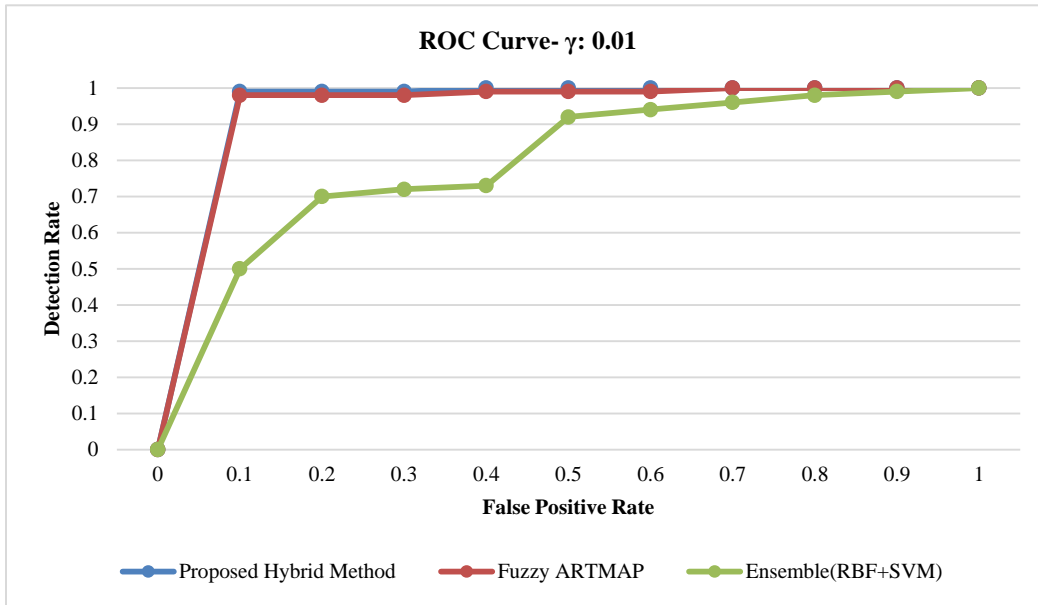
(c) Unknown assaults with parameter  $\gamma=1$  on ROC Curve

Fig. 6. Investigation and comparison of detection efficiency for unknown attacks using ROC curves.

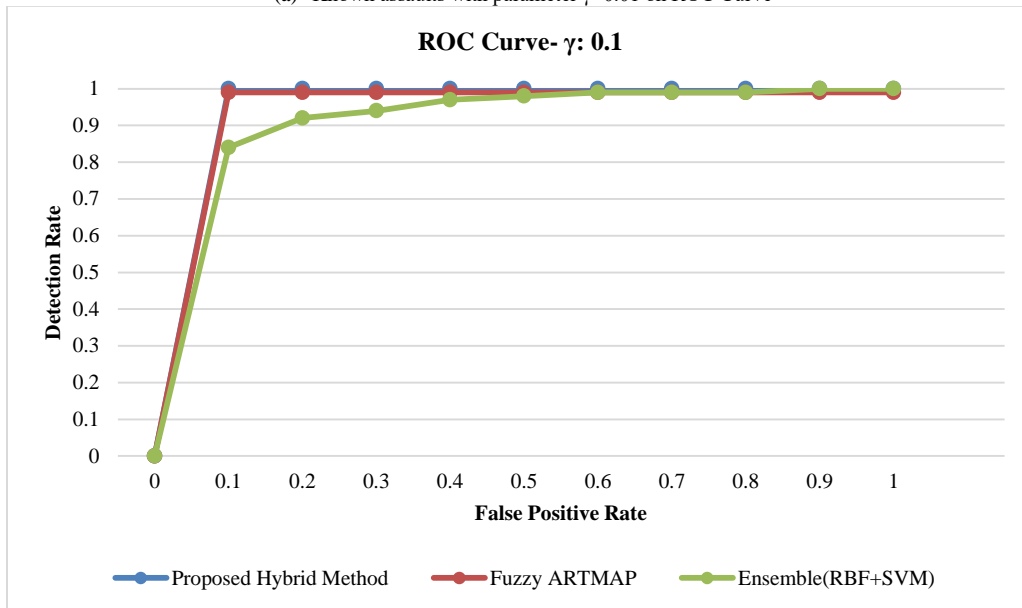
G. The Class 1 SVM Model's Detection Rate Approaches that of the Traditional Hybrid Method as

The parameter  $\gamma$  rises. This demonstrates that DT cannot influence the identification of unknown threats by enhancing Class 1 SVM performance. However, it has a major impact on hybrid intrusion detection models' ability to identify known attacks. Fig. 7, which varies the parameter  $c$  from 0.01 to 1, displays the ROC curves of the known attacks for the proposed technique and its comparisons. The figure shows that

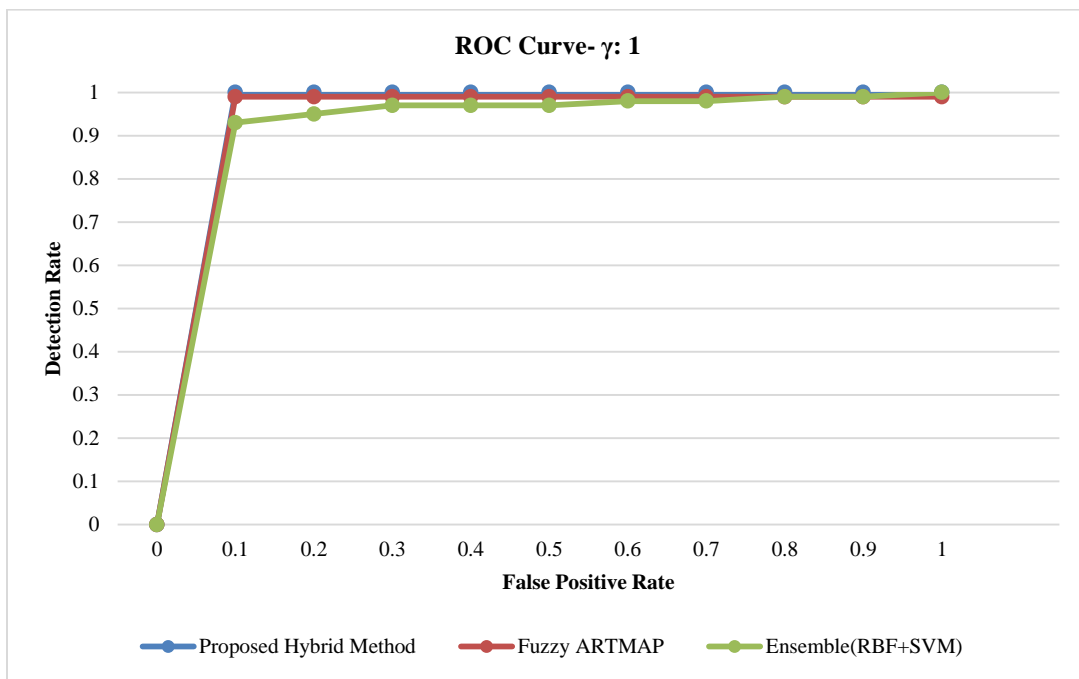
regardless of performance, both combined detection approaches from the SVM class 1 model have a detection rate of more than 99.1%. When  $\gamma$  is set to 1, the hybrid intrusion detection model's detection rate is somewhat greater than DT's, but its effectiveness is minimal. The class 1 SVM model's detection rate rises as  $c$  rises, as in the event of an unknown attack. However, in this instance, the class 1 SVM model-based anomaly detection approach is outperformed by the abuse detection method employing DT.



(a) Known assaults with parameter  $\gamma=0.01$  on ROC Curve



(b) Known assaults with parameter  $\gamma=0.1$  on ROC Curve



(c) known assaults with parameter  $\gamma=1$  on ROC Curve

Fig. 7. Investigation and comparison of detection efficiency for known attacks using ROC curves.

This paper introduces an innovative approach to intrusion detection in computer networks by combining abuse detection and anomaly detection techniques. This hybrid approach looks promising because it leverages the strengths of both methods and potentially leads to more effective network security. To select the algorithm, it is also attractive to use computational intelligence algorithms such as frog jumping algorithm, decision trees, support vector machines (SVM) and radial-based neural networks. These algorithms offer a variety of ways to handle intrusion detection, and it makes sense to choose them based on the specific needs of the system. It is also instructive about the execution speed and comparison between SVM and radial neural networks. It is noteworthy to observe that the use of RBF can significantly reduce model training time, as real-time intrusion detection often requires efficient algorithms. Overall, the paper provides a comprehensive overview of the proposed intrusion detection system, its methodology and performance metrics. It provides valuable insights into the potential benefits of combining exploit and anomaly detection techniques and provides a structured approach for future research in network security.

This paper presents a detailed evaluation of the proposed intrusion detection system, including results obtained from experiments using the NSL-KDD dataset. The evaluation includes various performance metrics and comparisons with other intrusion detection systems. Performance measures, detection rate, comparative analysis, training and test results, execution speed and ROC curves are among the evaluations discussed in this research. This paper concludes that the proposed hybrid intrusion detection system provides competitive detection rates and feature selection capabilities compared to other systems. This shows that the system performance can be improved by using RBF models instead of SVM in certain scenarios.

In general, the results and evaluations presented in the paper show the effectiveness of the proposed intrusion detection system in detecting known and unknown attacks. Using various performance metrics, comparative analysis and visual displays (such as ROC curves) provide a comprehensive assessment of system capabilities and tradeoffs. These results help to understand how computational intelligence algorithms can enhance network security.

## V. CONCLUSION

This paper provided a hybrid approach to network intrusion detection. This section will contrast the proposed mixing system's performance with that of comparable systems that employ feature selection algorithms and classification tools from various models, such as decision trees. The proposed mixing system utilizes multiple methods and models in its various components. SVM, Natural Bayes, and artificial neural networks have all been utilized for intrusion detection. The system test with KDD'99 or NSL-KDD data produced the findings that are shown in Table VII. The NSL-KDD dataset, which has the same number of characteristics as the original KDD'99 dataset, should be noted. Classifiers do not favor data with more repetitions since the extension records in the training set are destroyed during the compression process. The majority of detection rates, though, are higher on KDD'99 than NSL-KDD.

The hybrid approach suggested in this work has the best detection rate compared to other models tested on NSL-KDD data in prosperous years, as shown in Table VII. Meanwhile, the proposed models use fewer features than other models, which is a smaller number of features overall. As a result, it can be said that the suggested system offers a good mix of feature selection and classification techniques and that its

performance results are comparable to those of other systems that have been shown effective using KDD'99 and NSL-KDD data.

Additionally, the findings of the experiment demonstrated that while employing parallel SVMs, whose inputs are the leaves of a tree from smaller and more coherent data, resulting in a reduction in model training time, the model execution time relative to using RBF is increased. Instead of SVM, it is significantly longer (roughly 28 times the execution time of the model using RBF), which results in the model's inefficiency, particularly in high-speed networks; In the scenario where the execution errors and false alarm rates of the system are comparable in both models. It can be stated that using RBF instead of SVM has considerably increased the efficiency of the combined model because model training is an offline process, whereas testing and implementation are online processes.

This article examines the performance of the system in detecting known and unknown attacks. However, the system's effectiveness in detecting brand new and zero-day attacks has not been addressed. Future research can focus on developing methods to enhance the detection of previously unseen threats. We also pointed out in this research that using SVM instead of RBF can lead to a significant increase in execution time. Scalability is a critical concern for intrusion detection systems, especially in high-speed networks. Future research can explore ways to optimize the computational efficiency of the system without compromising the detection accuracy. In summary, future research in intrusion detection should focus on addressing these limitations and advancing the field by developing more robust systems. A more consistent and efficient focus can effectively detect a wide range of network intrusions while taking into account ethical and privacy considerations.

For those people who are eager for more research and new work in this field, it is suggested that in order to reduce false alarms in the anomaly detection stage, they should look for a solution to reduce the rate of false negative alarms in the abuse detection stage, because the outputs of the abuse detection stage which were categorized under the title of normal, are used as the input of the anomaly detection stage, and the anomaly detection model needs normal data as its input for better training and detection of violations from the normal pattern, and by reducing the negative rate of error in the detection stage abuse can achieve this.

#### REFERENCES

- [1] V. Kanimozhi and T. P. Jacob, "Artificial intelligence-based network intrusion detection with hyper-parameter optimization tuning on the realistic cyber dataset CSE-CIC-IDS2018 using cloud computing," in *2019 international conference on communication and signal processing (ICCSP)*, IEEE, 2019, pp. 33–36.
- [2] Y. Zhang, S. Wang, P. Phillips, and G. Ji, "Binary PSO with mutation operator for feature selection using decision tree applied to spam detection," *Knowl Based Syst*, vol. 64, pp. 22–31, 2014.
- [3] S. Khan, K. Kifayat, A. Kashif Bashir, A. Gurtov, and M. Hassan, "Intelligent intrusion detection system in smart grid using computational intelligence and machine learning," *Transactions on Emerging Telecommunications Technologies*, vol. 32, no. 6, p. e4062, 2021.
- [4] A. S. Eesa, Z. Orman, and A. M. A. Brifcani, "A novel feature-selection approach based on the cuttlefish optimization algorithm for intrusion detection systems," *Expert Syst Appl*, vol. 42, no. 5, pp. 2670–2679, 2015.
- [5] A. Shukla, S. Ahamad, G. N. Rao, A. J. Al-Asadi, A. Gupta, and M. Kumbhkar, "Artificial intelligence assisted IoT data intrusion detection," in *2021 4th International Conference on Computing and Communications Technologies (ICCT)*, IEEE, 2021, pp. 330–335.
- [6] J. Pirgazi and A. R. Khanteymooiri, "SFLA based gene selection approach for improving cancer classification accuracy," *AUT Journal of Modeling and Simulation*, vol. 47, no. 1, pp. 1–8, 2015.
- [7] Haoyan Zhang, Xudong Zhao, Huangqing Wang, Ben Niu, Ning Xu, Adaptive Tracking Control for Output-Constrained Switched MIMO Pure-Feedback Nonlinear Systems with Input Saturation, *Journal of systems science & complexity*, 36: 960–984, 2023.
- [8] Ning Xu, Zhongyu Chen, Ben Niu, and Xudong Zhao. Event-Triggered Distributed Consensus Tracking for Nonlinear Multi-Agent Systems: A Minimal Approximation Approach, *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, DOI: 10.1109/JETCAS.2023.3277544, 2023.
- [9] 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*, 360(14), 10728-10744.
- [10] G. Folino and P. Sabatino, "Ensemble based collaborative and distributed intrusion detection systems: A survey," *Journal of Network and Computer Applications*, vol. 66, pp. 1–16, 2016.
- [11] 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.
- [12] A. A. Aburumman and M. B. I. Reaz, "A novel SVM-kNN-PSO ensemble method for intrusion detection system," *Appl Soft Comput*, vol. 38, pp. 360–372, 2016.
- [13] A. Deshpande, "A Review on Intrusion Detection System using Artificial Intelligence Approach".
- [14] I. S. Thaseen and C. A. Kumar, "Intrusion detection model using fusion of chi-square feature selection and multi class SVM," *Journal of King Saud University-Computer and Information Sciences*, vol. 29, no. 4, pp. 462–472, 2017.
- [15] S. Shamshirband, M. Fathi, A. T. Chronopoulos, A. Montieri, F. Palumbo, and A. Pescapè, "Computational intelligence intrusion detection techniques in mobile cloud computing environments: Review, taxonomy, and open research issues," *Journal of Information Security and Applications*, vol. 55, p. 102582, 2020.
- [16] E. Khezri, E. Zeinali, and H. Sargolzaey, "A novel highway routing protocol in vehicular ad hoc networks using VMaSC-LTE and DBA-MAC protocols," *Wirel Commun Mob Comput*, vol. 2022, 2022.
- [17] S. R. Islam, W. Eberle, S. K. Ghafoor, A. Siraj, and M. Rogers, "Domain knowledge aided explainable artificial intelligence for intrusion detection and response," *arXiv preprint arXiv:1911.09853*, 2019.
- [18] T. Alladi, V. Kohli, V. Chamola, F. R. Yu, and M. Guizani, "Artificial intelligence (AI)-empowered intrusion detection architecture for the internet of vehicles," *IEEE Wirel Commun*, vol. 28, no. 3, pp. 144–149, 2021.
- [19] V. Kanimozhi and T. P. Jacob, "Artificial Intelligence outflanks all other machine learning classifiers in Network Intrusion Detection System on the realistic cyber dataset CSE-CIC-IDS2018 using cloud computing," *ICT Express*, vol. 7, no. 3, pp. 366–370, 2021.
- [20] S. Alzahrani and L. Hong, "Detection of distributed denial of service (DDoS) attacks using artificial intelligence on cloud," in *2018 IEEE World Congress on Services (SERVICES)*, IEEE, 2018, pp. 35–36.
- [21] Wang, Z., Jin, Z., Yang, Z., Zhao, W., & Trik, M. (2023). Increasing efficiency for routing in Internet of Things using Binary Gray Wolf Optimization and fuzzy logic. *Journal of King Saud University-Computer and Information Sciences*, 101732.
- [22] Haoyu Zhang, Quan Zou, Ying Ju, Chenggang Song, Dong Chen. Distance-based Support Vector Machine to Predict DNA N6-methyladine Modification. *Current Bioinformatics*. 2022, 17(5): 473-482.

- [23] 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.
- [24] E. Khezri and E. Zeinali, "A review on highway routing protocols in vehicular ad hoc networks," *SN Comput Sci*, vol. 2, pp. 1–22, 2021.
- [25] Yan, S., Gu, Z., Park, J.H., and Xie, X., 2023. A delay-kernel-dependent approach to saturated control of linear systems with mixed delays. *Automatica*, 152, p. 110984.
- [26] A. Branitskiy and I. Kotenko, "Hybridization of computational intelligence methods for attack detection in computer networks," *J Comput Sci*, vol. 23, pp. 145–156, 2017.
- [27] I. F. Kilincer, F. Ertam, and A. Sengur, "Machine learning methods for cyber security intrusion detection: Datasets and comparative study," *Computer Networks*, vol. 188, p. 107840, 2021.
- [28] T. C. Truong, I. Zelinka, J. Plucar, M. Čandík, and V. Šulc, "Artificial intelligence and cybersecurity: Past, presence, and future," in *Artificial intelligence and evolutionary computations in engineering systems*, Springer, 2020, pp. 351–363.
- [29] S. Zhao, S. Li, L. Qi, and L. Da Xu, "Computational intelligence enabled cybersecurity for the internet of things," *IEEE Trans Emerg Top Comput Intell*, vol. 4, no. 5, pp. 666–674, 2020.
- [30] M. Samiei, A. Hassani, S. Sarspy, I. E. Komari, M. Trik, and F. Hassanpour, "Classification of skin cancer stages using a AHP fuzzy technique within the context of big data healthcare," *J Cancer Res Clin Oncol*, pp. 1–15, 2023.
- [31] J. Sun, Y. Zhang, and M. Trik, "PBPHS: a profile-based predictive handover strategy for 5G networks," *Cybern Syst*, pp. 1–22, 2022.
- [32] R. Patgiri, U. Varshney, T. Akutota, and R. Kunde, "An investigation on intrusion detection system using machine learning," in *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, IEEE, 2018, pp. 1684–1691.
- [33] M. Trik, H. Akhavan, A. M. Bidgoli, A. M. N. G. Molk, H. Vashani, and S. P. Mozaffari, "A new adaptive selection strategy for reducing latency in networks on chip," *Integration*, vol. 89, pp. 9–24, 2023.
- [34] M. Trik, A. M. N. G. Molk, F. Ghasemi, and P. Pouryeganeh, "A hybrid selection strategy based on traffic analysis for improving performance in networks on chip," *J Sens*, vol. 2022, 2022.
- [35] N. Shone, T. N. Ngoc, V. D. Phai, and Q. Shi, "A deep learning approach to network intrusion detection," *IEEE Trans Emerg Top Comput Intell*, vol. 2, no. 1, pp. 41–50, 2018.
- [36] Somanath, S., Wakkary, R., Ettehad, O., Lin, H., Behzad, A., Eshpeter, J., & Oogjes, D. (2022). Exploring the composite intentionality of 3D printers and makers in digital fabrication. *International Journal of Design*, 16(3), 77-95.
- [37] M. Trik, S. P. Mozaffari, and A. M. Bidgoli, "Providing an adaptive routing along with a hybrid selection strategy to increase efficiency in NoC-based neuromorphic systems," *Comput Intell Neurosci*, vol. 2021, 2021.
- [38] D. Mokhlesi Ghanevati, E. Khorami, B. Boukani, and M. Trik, "Improve replica placement in content distribution networks with hybrid technique," *Journal of Advances in Computer Research*, vol. 11, no. 1, pp. 87–99, 2020.
- [39] S. Hanif, T. Ilyas, and M. Zeeshan, "Intrusion detection in IoT using artificial neural networks on UNSW-15 dataset," in *2019 IEEE 16th international conference on smart cities: improving quality of life using ICT & IoT and AI (HONET-ICT)*, IEEE, 2019, pp. 152–156.
- [40] 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.
- [41] E. Khezri, E. Zeinali, and H. Sargolzaey, "SGHRP: Secure Greedy Highway Routing Protocol with authentication and increased privacy in vehicular ad hoc networks," *PLoS One*, vol. 18, no. 4, p. e0282031, 2023.
- [42] Golrou, A., Rafiei, N., & Sabouri, M. Wheelchair Controlling by eye movements using EOG based Human Machine Interface and Artificial Neural Network. *International Journal of Computer Applications*, 975, 8887.
- [43] I. H. Abdulqadder, S. Zhou, D. Zou, I. T. Aziz, and S. M. A. Akber, "Multi-layered intrusion detection and prevention in the SDN/NFV enabled cloud of 5G networks using AI-based defense mechanisms," *Computer Networks*, vol. 179, p. 107364, 2020.

# Autism Diagnosis using Linear and Nonlinear Analysis of Resting-State EEG and Self-Organizing Map

Jie Xu<sup>1</sup>, Wenxiao Yang<sup>2\*</sup>

Anyang Normal University, Anyang 455000, Henan, China<sup>1</sup>

Graduate Affairs Office, Shanghai University of Medicine & Health Sciences, Shanghai 201318, Shanghai, China<sup>2</sup>

**Abstract**—The prevalence of autism has increased dramatically in recent years and many people around the world are facing this difficult condition. There is a need to develop an objective method to diagnose autism. Various analysis methods have been used to classify the EEG signals of people with autism, from linear methods in the time and frequency domain to nonlinear methods based on chaos theory. However, there is still no consensus on which method of EEG signal analysis can provide us with the best diagnostic accuracy and valid biomarkers for autism diagnosis. Therefore, in this study, we evaluate different feature extraction methods from EEG signals to diagnose autism from healthy individuals. For this purpose, EEG analysis was performed in time, time-frequency, frequency and nonlinear domains. Furthermore, the self-organizing map (SOM) method was used to classify features extracted from autistic and normal EEG. The data used in this study were recorded by the research team from 24 children with autism and 24 normal children. The accuracies of 92.31, 93.57, 95.63 and 97.10% were achieved through time and morphological, frequency, time-frequency and nonlinear analyzes, respectively. Indeed, the findings showed that nonlinear analysis could yield the best classification results (accuracy = 97.10%, sensitivity = 98.80% and specificity = 97.02%) in the EEG discrimination of autistic children from typical children through the SOM neural network.

**Keywords**—Autism; EEG; linear analysis; nonlinear analysis; neural network

## I. INTRODUCTION

Autism is a severe psychiatric disease in which patients have serious problems in executive functions, social relations, cognition, normal behaviors and daily activities [1]. Its prevalence has grown in recent years drastically, and many people around the world are facing this difficult condition [2-4]. However, psychiatrists and psychologists do not deal with the definitive state of this disorder, but they have to deal with autism spectrum disorder with different biological and behavioral symptoms [5, 6]. This causes doctors to choose different screening tools to diagnose patients with autism and therapists to choose different treatment approaches for each patient [7]. But most of the existing screening and diagnostic tools are subjective, and various types of research suggest the need to develop an objective method to diagnose autism [8]. For example, a review article highlighted various biomarkers, such as hormones, to develop reliable, objective methods for diagnosing autism [9]. A systematic review focused on the

application of artificial intelligence in autism screening and diagnosis through validated questionnaire-based data such as the Autism Diagnostic Interview-Revised (ADI-R) and the Autism Diagnostic Observation Schedule (ADOS) [10]. Another systematic review showed that a combination of eye-tracking technology and machine learning could be taken into account as a suitable approach for objective and early diagnosis of autism [11]. Lai et al. proposed an objective method for autism diagnosis based on automatic retinal image analysis and machine learning and reported a good accuracy of 97.4% for this purpose [12]. Zhao et al. proposed an automatic objective system based on the analysis of the movements of patients during a motor task and machine learning algorithms. They reported an accuracy of 88.37% for autism diagnosis [13]. Therefore, as we can see from the literature, many researchers around the world have tried to develop objective methods of autism diagnosis through various psychological, biological and physiological data.

In the meantime, electroencephalogram (EEG) is one of the electrophysiological data that has received much attention from researchers and has been analyzed in various ways to develop objective methods for diagnosing autism [14]. EEG indicates the pattern of neural electrical activity in different areas of the brain, providing brilliant information about brain function in various healthy and unhealthy conditions [15, 16]. Therefore, EEG signals have been targeted by computational neuroscientists and biomedical engineers for various biomedical applications [17-23]. So far, various EEG biomarkers have been introduced to diagnose psychiatric and neurological disorders [24, 25]. Bosl et al. used a combination of nonlinear EEG analysis and different machine learning techniques, achieving a 95% accuracy in screening pediatric populations at risk for autism [26]. Haputhanthri et al. extracted statistical features from wavelet analysis applied to EEG signals of children with autism, achieving a diagnostic accuracy of 93% [27]. Ahmadlou et al. proposed a fuzzy synchronization likelihood wavelet approach, achieving an EEG classification accuracy of 95.5% for autism diagnosis [28]. Pham et al. applied the higher-order spectra bispectrum method to EEG signals and achieved a high accuracy of 98.7% using a probabilistic neural network for autism diagnosis [29]. Baygin et al. utilized a combined deep lightweight feature extraction method based on one-dimensional local binary patterns and deep features of the spectrogram images generated by the short-time Fourier transform. The 10-fold cross-

validation algorithm showed an ability to identify children with autism with a support vector machine with 96.44% accuracy [30]. Alotaibi and Maharatna proposed an EEG classification system based on functional connectivity features conceptualized by graph theory and cubic support vector machine, achieving an accuracy of 95.8% for autism diagnosis [31]. Radhakrishnan et al. evaluated deep learning models for the diagnosis of autism from EEG signals, reporting an average accuracy of 81% using their methodology [32].

As can be seen in the literature, various analysis methods have been used to classify the EEG signals of people with autism, from linear methods in the time and frequency domain to nonlinear methods based on chaos theory. However, there is still no consensus on which method of EEG signal analysis can provide us with the best diagnostic accuracy and provide valid biomarkers for autism diagnosis. Therefore, in this study, we are going to evaluate different feature extraction methods from EEG signals in order to diagnose autism from healthy individuals. For this purpose, EEG analysis in time, time-frequency, frequency, and nonlinear domains were performed through various analysis techniques. This paper is organized as follows. Section II provides the procedure proposed in the current study. Section III reports the experimental results. Section IV discusses the obtained results and makes a conclusion.

## II. METHODS

In this section, various analysis methods applied to EEG signals for autism diagnosis were described. Fig. 1 shows the framework adopted in this study.

### A. Time and Morphological Analysis

EEG signals have specific temporal characteristics that may be affected by different neuropathologies. As a result, according to these temporal characteristics, it is possible to

extract features based on the signal waveform over time, which can be used to distinguish between the two classes of autism and normal. These features are simple and require very little computation. The advantage of such features is increasing the speed of the designed system and the possibility of using it in real-time [33-35]. The following features in this category were calculated from EEG signals.

$$\text{Absolute amplitude} = \max|s(t)| \quad (1)$$

$$\text{Peak-to-Peak} = S_{max} - S_{min} \quad (2)$$

$$\text{Negative Area} = \sum_t 0.5(s(t) - |s(t)|) \quad (3)$$

$$\text{Positive Area} = \sum_t 0.5(s(t) + |s(t)|) \quad (4)$$

$$\text{Total Absolute Area} = |\text{Negative Area}| + \text{Positive Area} \quad (5)$$

$$\text{Zero Crossing} = \sum_t \delta_s(t) \cdot \delta_s(t) = \begin{cases} 1 & s(t) \times s(t-1) < 0 \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

$$\text{Mean} = \frac{1}{n} \sum_{t=1}^n s_t \quad (7)$$

$$\text{Standard deviation} = \left( \frac{1}{n} \sum_{t=1}^n (s_t - \bar{s})^2 \right)^{\frac{1}{2}} \quad (8)$$

$$\text{Energy} = \sum_t |s(t)|^2 \quad (9)$$

$$\text{Skewness} = \frac{E[(s-\mu)^3]}{\sigma^3} \quad (10)$$

$$\text{Kurtosis} = \frac{E[(s-\mu)^4]}{\sigma^4} \quad (11)$$

where  $s(t)$  is the time series under analysis,  $n$  is the number of data points,  $\mu$  is the mean of the time series,  $\sigma$  is the standard deviation, and  $E$  denotes the expectation operator.

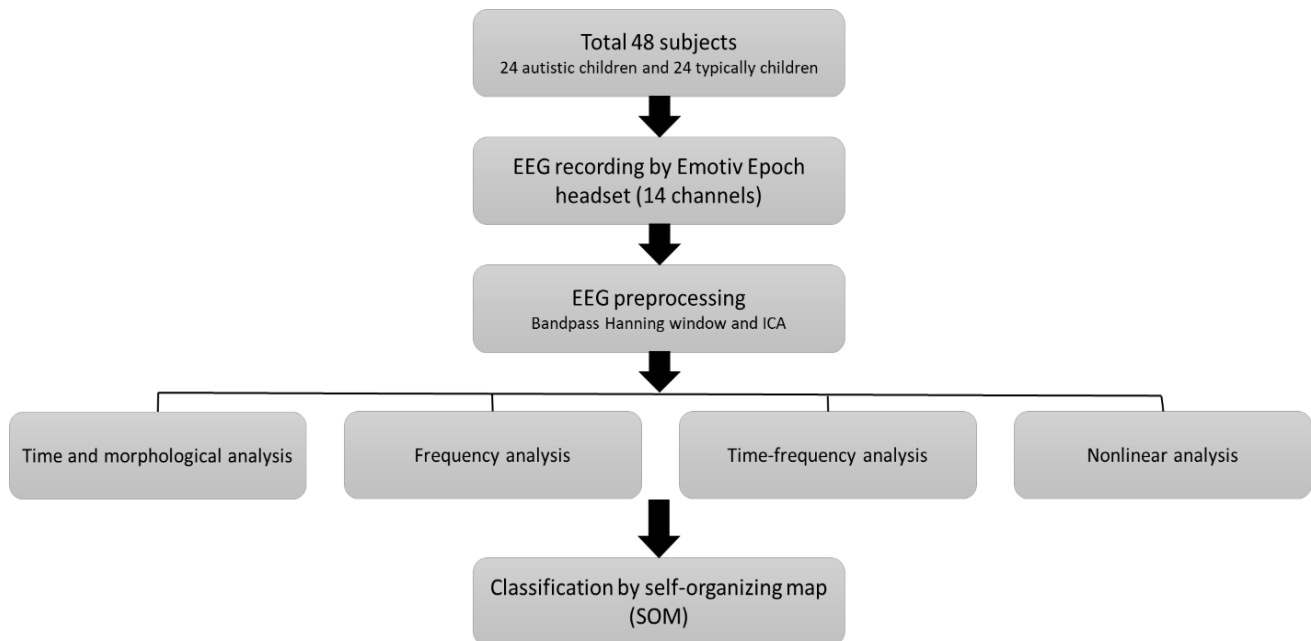


Fig. 1. Adopted framework in this study for EEG classification of autistic children and typically children.



### B. Frequency Analysis

Frequency features actually represent the rate of change in the signal. Methods such as the Fourier transform are used to convert the signal from the time domain to the frequency domain. Here, the Welch method was utilized to extract EEG sub-bands, including delta (1-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), sigma (12-16 Hz), beta (16-30 Hz) and gamma (30-45 Hz). This method is Fourier transform-based algorithm to estimate the power spectral density. After signal sub-bands extraction and power spectrum density estimation, we calculated mean, standard deviation, skewness, kurtosis, absolute power and relative power from each sub-band as frequency features.

### C. Time-Frequency Analysis

We utilized this type of analysis to assess signals in the time and frequency domains concurrently. For this purpose, wavelet transform was used, which provides a time-frequency representation of EEG signals with good frequency and time localization. This technique decomposes time series into shifted and scaled versions of the basic wavelet function. The wavelet function can be written as:

$$\psi_{a,b}(t) = 2^{a/2}\psi(2^{a/2}(t - b)) \quad (12)$$

where  $\psi(t)$  denotes the wavelet function,  $a$  is the scale parameter, and  $b$  is the shift parameter. The discrete version of this algorithm decomposes EEG signals into high- and low-frequency components at each level, known as detail and approximation coefficients [19]. In the current work, the Haar wavelet was employed to represent the time-frequency sub-components of the signals. After calculating the detail and approximation coefficients, we calculated the mean, standard deviation, variance and entropy as time-frequency features.

### D. Nonlinear Analysis

Due to the nonlinear characteristics of EEG, nonlinear analyzes may reveal more details about the neuropathological mechanisms involved in autism. In the current study, we tried to calculate various nonlinear features for the signals, including large Lyapunov exponent, Lempel-Ziv measure, approximate entropy, sample entropy, Higuchi fractal dimension, and detrended fluctuation analysis. In this subsection, the mathematical notation of these nonlinear features was explained.

1) *Large lyapunov exponent*: This feature is a chaotic concept to assess the trajectory divergence in dynamical systems. Lyapunov exponent determines the exponential divergence rate between the two adjacent trajectories. The mathematical notation of this exponent can be written as:

$$\lambda = \frac{1}{n} \ln \left( \frac{d_n}{d_0} \right) \quad (13)$$

where  $d_n$  and  $d_0$  denote the divergence/distance between sequential data points in the  $n^{\text{th}}$  and initial times, respectively.

2) *Lempel-Ziv measure*: It is a complex feature to estimate new paradigms in EEG signals. This method converts a signal to a binary one by median thresholding and scans the binary signal for new subsequences in sequential symbols [36].

$$LZ = \frac{\log_2(n) \cdot c(n)}{n} \quad (14)$$

where  $n$  is the number of data points and  $c(n)$  denotes the number of new subsequences.

3) *Approximate entropy*: It is a measure to estimate the randomness of EEG fluctuations over time.

$$ApEn(m, r, N) = \ln \left[ \frac{C_m(r)}{C_{m+1}(r)} \right] \quad (15)$$

where  $C_m(r)$  denotes the repeating paradigms of length  $m$  in a signal of  $N$  data points according to the similarity index  $r$ . In this work, we set  $m = 2$  and  $r = 0.2$  standard deviation of EEG signals [37].

4) *Sample entropy*: Sample entropy: It is a modified algorithm of approximate entropy that reduces the self-matching bias in the entropy calculation [38]. This algorithm depends on the length of the data and yields relatively consistent results in various conditions.

$$SampEn(m, r, N) = -\ln \left( \frac{A^m(r)}{B^m(r)} \right) \quad (16)$$

where  $r$ ,  $m$  and  $N$  indicate tolerance, embedding dimension and the number of samples, respectively.  $B^m(r)$  represents the probability that two series of data samples of length  $m$  have a distance smaller than the tolerance  $r$ , and  $A^m(r)$  indicates a similar probability for two series of data samples of length  $m+1$ .

5) *Higuchi fractal dimension*: Consider a time series  $S(N) = S(1), S(2), \dots, S(N)$  as an input. Higuchi algorithm builds a new time series based on  $S(N)$  as follows:

$$S_m^k = \left\{ S(m), S(m+k), S(m+2k) \dots S \left( m + \left\lfloor \frac{N-m}{k} \right\rfloor k \right) \right\}. \text{ for } m = 1, 2, 3 \dots k \quad (17)$$

where  $m$  is the first sample of the time series and  $\left\lfloor \frac{N-m}{k} \right\rfloor$  represents the integer part of the series. Length  $L_m(k)$  for  $S_m^k$  is obtained by:

$$L_m(k) = \frac{\sum_{i=1}^k |S(m+ik) - S(m+(i-1)k)| (N-1)}{\left\lfloor \frac{N-m}{k} \right\rfloor k} \quad (18)$$

where  $N$  represents the number of total samples in the time series and  $\frac{(N-1)}{\left\lfloor \frac{N-m}{k} \right\rfloor k}$  represents the normalization coefficient. The total mean length,  $L(k)$ , is calculated for  $k_1$  to  $k_{\max}$  for all  $k$ .

6) *Detrended fluctuation analysis*: The DFA criterion is used to reveal the correlation of the time series with itself in the long-term time range [39]. To calculate the DFA in the time series, it must first be aggregated according to the following relationship.

$$y(k) = \sum_{i=1}^k [x(i) - x_{\text{average}}] \quad (19)$$

Then  $y(k)$  is divided into equal segments of length  $n$ . One line fits each segment. This line is denoted by  $y_n(k)$ , and  $y_n(k)$  is subtracted from  $y(k)$ .

$$F(n) = \sqrt{\frac{1}{N} \sum_{k=1}^N [y(k) - y_n(k)]^2} \quad (20)$$

To obtain the DFA, one must obtain the equivalent of  $F(n)$  for a suitable number of  $n$ . Then its graph is drawn in a logarithmic scale, and the slope of the scaling area is introduced as DFA.

### E. Classification

In the current work, the self-organizing map (SOM) method was used to classify features extracted from autistic and normal EEG. This method is a popular unsupervised neural network with many applications in prediction, classification and clustering problems [40]. In this algorithm, a vector of weights is defined for every neuron  $i$ . The dimension of this vector is equal to the dimension of the input data. Firstly, a winner neuron is specified by the following equation:

$$i^*(t) = \operatorname{argmin}\{g_i(x(t))\} \quad . \quad g_i(x(t)) = \|x(t) - w_i(t)\| \quad (21)$$

where  $w_i(t)$  is the weight vector that must be updated based on the following equation:

$$\Delta w_i(t) = \alpha(t)h(i^*.i;t)[x(t) - w_i(t)] \quad (22)$$

where  $h(\cdot)$  denotes the neighborhood function with the following definition:

$$h(i^*.i;t) = \exp(-\|r_i(t) - r_{i^*}(t)\|^2 / \sigma^2(t)) \quad (23)$$

$\|r_i(t) - r_{i^*}(t)\|$  defines the distance between  $i$  and  $i^*$ ,  $\sigma(t)$  is the neighborhood radius, and  $\alpha(t)$  is the learning rate parameter.

To be classified by SOM, the output neurons must be labeled. After training the SOM, a winner neuron is devoted to each training vector. Then, the label of each training vector is determined. Eventually, the label of the winner neuron is defined based on the most frequent class labels of the training vectors. In this work, the initial neighborhood parameter was defined as 3, which was reduced to 1 after 100 iterations. Moreover,  $\alpha(t)$  was set at 0.8.

### F. EEG Data

The data used in this study were recorded by the research team from 24 children with autism and 24 normal children. Participants ranged in age from 4-9 years, and all patients received a diagnosis of autism based on DSM-5 diagnostic criteria by experienced clinicians. The patient enrollment was administered in a psychiatric clinic. The research project was done in accordance with the principles of the Declaration of Helsinki (1996) and the current Good Clinical Practice guidelines. The goal and an overview of the project were characterized by the participants and their parents during the initial contact. For those who agreed to participate, all the necessary information was provided prior to signing written informed consent. Information about the subjects was utilized anonymously and for the purpose of the study.

EEG was recorded for 10-18 minutes for each participant in one session. Given the difficulties of working with autistic patients and the difficulties of recording EEG from these patients in the awake state, the Emotiv Epoch headset device

was employed in this research. Since the Emotiv Epoch headset is a wireless EEG device, the signal recording was conducted in autistic patients more easily. This EEG device uses a Bluetooth module for wireless communication. The Emotiv Epoch headset and Software Development Kit include 14 electrodes (AF3, AF4, F7, F8, F3, F4, FC5, FC6, T7, T8, P7, P8, O1, O2 based on 10-20 international system) along with DRL/CMS references at P4/P3 locations. The sampling rate in this device is 128 Hz. The impedance of the electrode is reduced through saline liquid and alcohol pads. Emotive software was utilized to record EEGs and convert their format to MATLAB format.

After signal recording, in the signal pre-processing step, a band-pass Hanning window with a finite duration and frequency range of 1-45 Hz was applied to the EEGs through MATLAB software. Furthermore, electrode interpolation was done through adjacent channels for low-quality electrodes. EEGs were re-referenced to the common average and then were decomposed via independent component analysis. Components with motion and muscle artifacts were identified and were then eliminated according to time courses and frequency scalp maps. The cleaned components were reconstructed, and a 50-second cleaned EEG signal was prepared for each participant.

## III. EXPERIMENTAL RESULTS

After data conditioning, all mentioned analyzes were applied to EEG signals and different described features were extracted in both typically and autism groups. Fig. 2 shows an example of an EEG signal recorded from a child with autism before pre-processing. Fig. 3 shows the time-frequency representation of two channels, P7 and P8, for normal and autistic children obtained from wavelet analysis. As shown in this figure, there are clear differences in the frequency content of the EEG signals of normal and autistic children over time. In addition, Fig. 4 to 6 show the nonlinear features (i.e., sample entropy, DFA and Lempel-Ziv measure) extracted from EEG signals of normal and autistic children in the O1 channel. These graphs show that there is a clear difference between the nonlinear dynamics of the EEG signals of the two groups. The noteworthy point is that the values of nonlinear features in the normal group were generally higher than in the autism group.

In the next step, we tried to classify different features extracted from EEGs through various analyzes. In this step, the leave-one-subject-out cross-validation method was utilized to validate the efficiency of every analysis method as well as the performance of the SOM classifier for autism diagnosis. In this cross-validation method, a subject was left out to test, and the rest of the subjects were utilized to train the SOM. As a result, after implementing the 48 tests, the average accuracy was calculated over all the obtained accuracies. Specificities, sensitivities and averaged classification accuracies for each type of analysis are depicted in Table I. Accuracies of 92.31, 93.57, 95.63 and 97.10% were achieved through time and morphological, frequency, time-frequency and nonlinear analyzes, respectively. Indeed, the findings showed that nonlinear analysis could yield the best classification results (accuracy = 97.10%, sensitivity = 98.80% and specificity =

97.02%) in the EEG discrimination of autistic children from typical children through the SOM neural network.

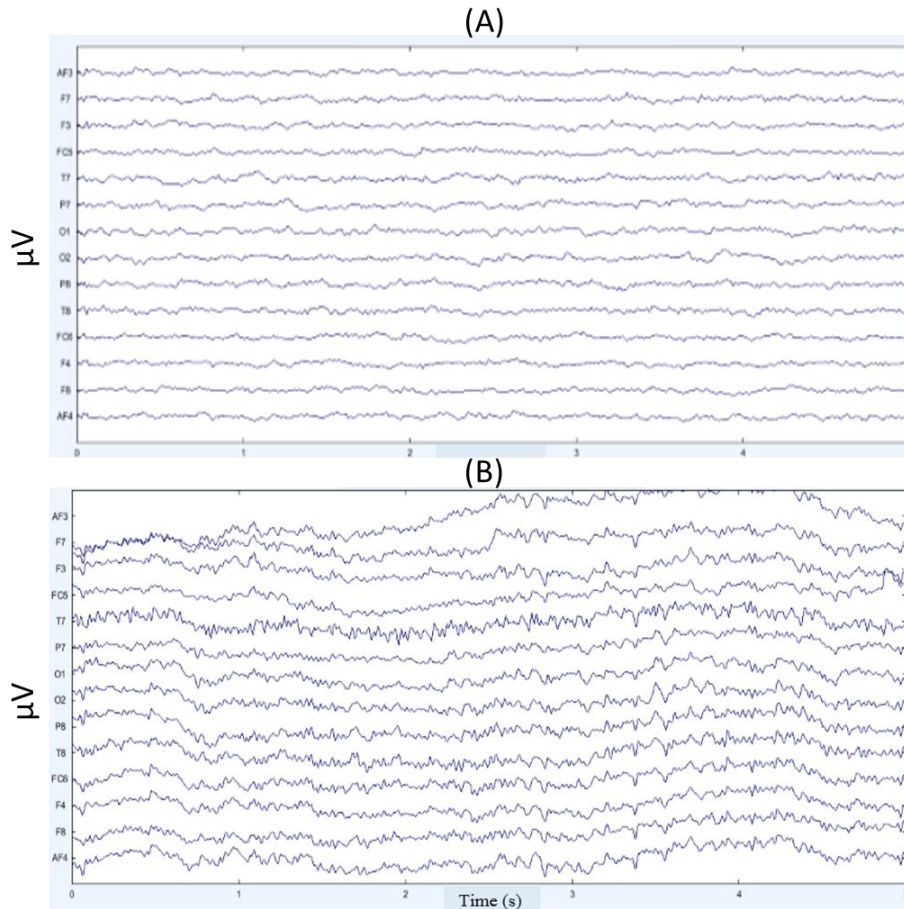


Fig. 2. An example of EEG signals recorded from (A) a healthy child and (B) a child with autism.

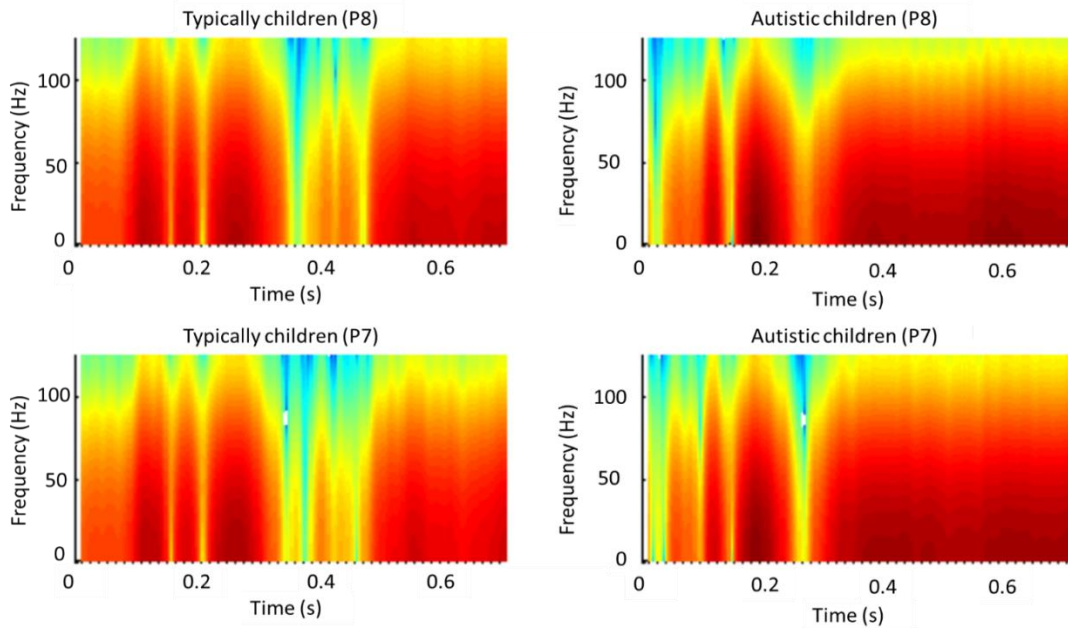


Fig. 3. Time-frequency representation of two channels, P7 and P8, for normal and autistic children.

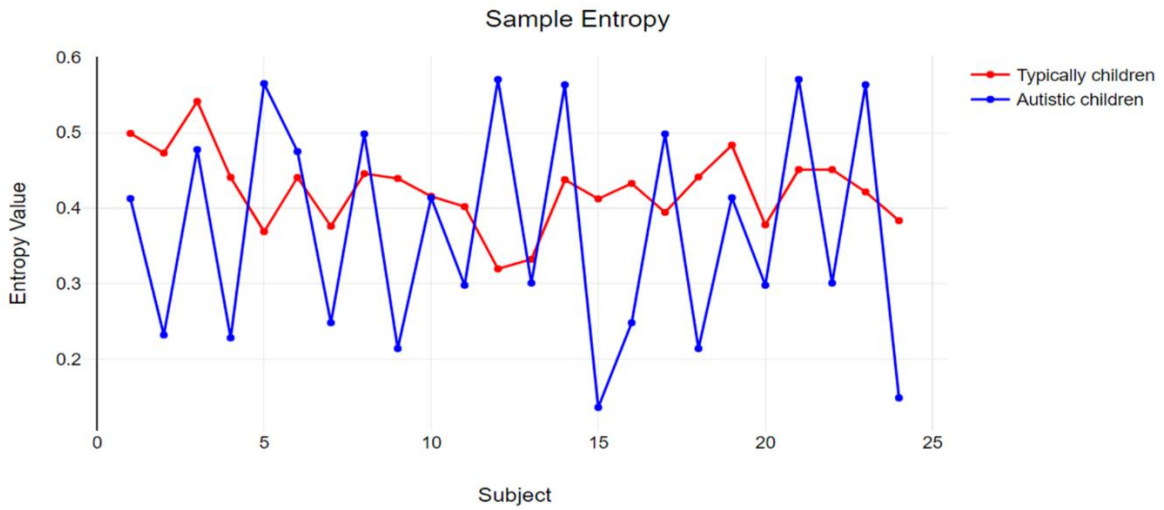


Fig. 4. Calculated sample entropy at the O1 channel for typical and autistic children.

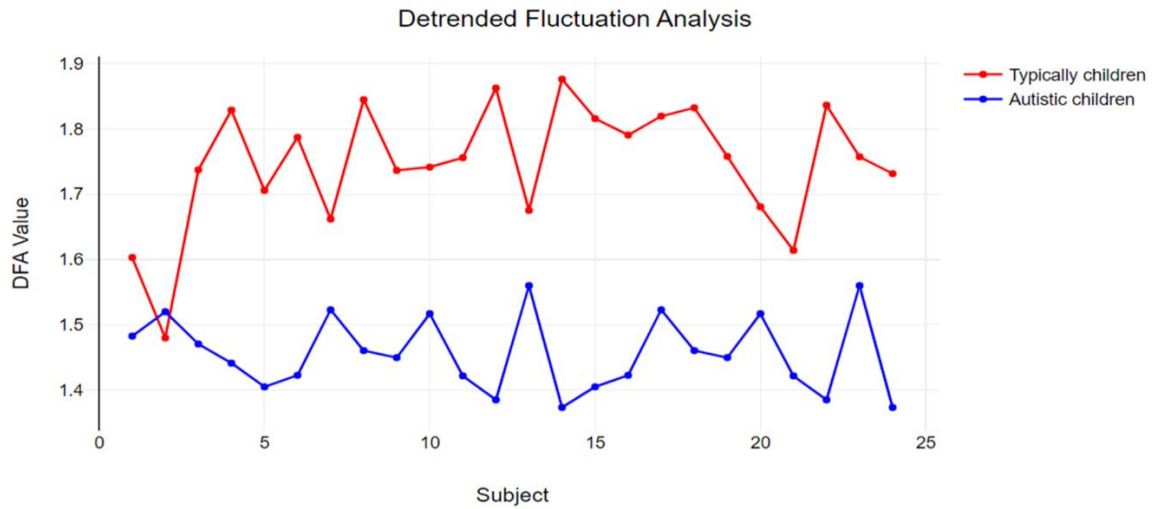


Fig. 5. Detrended fluctuation analysis at O1 channel for typically and autistic children.

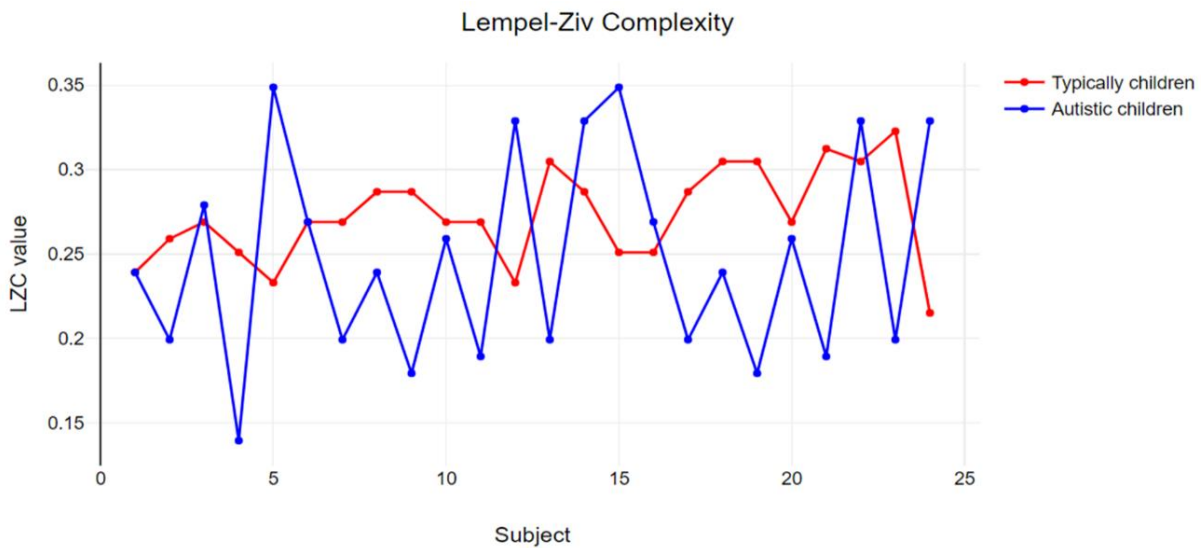


Fig. 6. Calculated Lempel-Ziv measure at O1 channel for typically and autistic children.

TABLE I. ACCURACY, SENSITIVITY AND SPECIFICITY FOR DIFFERENT ANALYZES USING SOM CLASSIFIER

Type of analysis	Accuracy (train) (%)	Accuracy (test) (%)	Sensitivity (%)	Specificity (%)
Time and morphological analysis	94.78 ± 4.25	92.31 ± 3.59	90.36	93.39
Frequency analysis	96.13 ± 3.36	93.57 ± 4.61	93.00	94.69
Time-frequency analysis	97.71 ± 2.06	95.63 ± 2.48	95.89	94.26
Nonlinear analysis	99.54 ± 2.14	97.10 ± 1.95	98.80	97.02

#### IV. DISCUSSION AND CONCLUSION

Autism is a neurodevelopmental condition that is related to different neural and neurotransmission impairments in various brain areas. These functional abnormalities of the brain are supposed to play a critical role in the neuropathology of autism [41, 42]. Therefore, the search for a reliable biomarker through EEG analysis is a hot topic in autism research. In the present study, we aimed to explore the different types of EEG analysis for feature extraction for autism diagnosis. For this purpose, time and morphological, time-frequency, frequency and nonlinear features were extracted from EEG signals of typical and autistic children at resting-state. The obtained findings revealed that nonlinear features achieved the best classification results for autism diagnosis. Indeed, the proposed nonlinear features integrated with the SOM classifier yielded an average accuracy of 97.10% in detecting autism cases, which is a good result for improving research achievements in this field. This type of quantitative analysis is more consistent with the nonlinear and chaotic properties of brain signals. This finding is in line with previous studies [43-47]. In other words, based on the findings of the present study, it is recommended that future studies focus more on various nonlinear EEG analysis methods and their optimization in order to diagnose autism. Since none of the previous EEG studies on autism have conducted a comparative study between different linear and nonlinear analysis methods, the findings of the present study as the first example in this field can be a roadmap for future research. However, this study, like many other studies, has limitations. The limited sample size is one of the important

limitations of the current research, which reduces the generalizability of the obtained findings. In this study, only five nonlinear analysis methods were investigated, while there are many more nonlinear methods and future studies should investigate different methods. In addition, we only analyzed resting-state EEG, while other recording protocols may have helped to improve the results.

Table II summarizes the characteristics of the studies on autism diagnosis using EEG analysis and machine learning. As shown in this table, previous studies used different methods for feature extraction from EEG signals, from linear frequency analysis to various nonlinear approaches such as recurrence quantification analysis and fractal dimension. The obtained results showed that future studies should work on the nonlinear dynamics of EEG signals and the combination and optimization of nonlinear features for autism diagnosis. Support vector machine (SVM) is the most frequently used classifier in these works to classify EEG features. Moreover, some studies used neural networks such as radial basis function and probabilistic neural networks for this purpose. Most studies have achieved an autism classification accuracy of 90% or higher, which shows the high potential of this approach for the objective diagnosis of autism. However, most of these studies suffer from important limitations that reduce their generalizability. Small datasets, complex implementation processes and low accuracy are some of these limitations. In addition, the results obtained in the present study compared to previous works show that the nonlinear approach adopted along with the SOM classification has a very good ability to diagnose autism.

TABLE II. CHARACTERISTICS OF THE STUDIES ON AUTISM DIAGNOSIS USING EEG ANALYSIS AND MACHINE LEARNING

Study	Population	Feature Extraction	Classifier	Results
Ahmadlou et al. (2010) [48]	Nine autistic and eight non-autistic children	Higuchi and Katz fractal dimension, wavelet-chaos neural network	Radial basis function (RBF)	Accuracy = 90%
Bosl et al. (2011) [49]	46 infants at high risk for autism and 33 healthy controls	Modified multiscale entropy	Support vector machine (SVM)	Accuracy = 90%
Ahmadlou et al. (2012) [28]	Nine autistic and nine healthy children	Fuzzy synchronization likelihood	Enhanced probabilistic neural network	Accuracy = 95.5%
Sheikhani et al. (2012) [50]	17 autistic children and 11 healthy children	Short-time Fourier transform	KNN	Accuracy = 96.4%
Jamal et al. (2014) [51]	12 subjects in each autism and normal group	Brain connectivity	Linear discriminant analysis (LDA) and SVM	Accuracy = 94.7%
Eldridge et al. (2014) [52]	19 autistic children and 30 healthy children	Modified multiscale entropy	SVM, Logistic regression, Naïve Bayes	Accuracy = 79%
Bosl et al. (2018) [26]	99 infants with an older sibling diagnosed with autism	Wavelet analysis, Sample entropy, DFA, Recurrence quantitative analysis	SVM	Accuracy = 95%
Heunis et al. (2018) [45]	Seven autistic children and seven non-autistic children	Recurrence quantitative analysis	SVM	Accuracy = 92.9%
Kang et al. (2018) [53]	52 autistic children and 52 non-autistic children	Fast Fourier transform	SVM	Accuracy = 91.38%

Haputhanthri et al. (2019) [27]	Ten autistic children and five non-autistic children	Wavelet analysis	Logistic regression, SVM, Naïve Bayes, Random forest	Accuracy = 93%
Pham et al. (2020) [29]	40 autistic children and 37 healthy children	higher-order spectra (HOS) bispectrum	LDA, SVM, k-nearest neighbor (KNN), probabilistic neural network (PNN)	Accuracy = 98.7%
Baygin et al. (2021) [30]	61 autistic subjects and 61 healthy subjects	one-dimensional local binary pattern and deep features of the spectrogram images	SVM	Accuracy = 96.44%
Alotaibi et al. (2021) [31]	12 autistic children and 12 healthy children	Brain connectivity	SVM	Accuracy = 95.8%
Our proposed approach	24 autistic children and 24 healthy children	Time, time-frequency, frequency, and nonlinear analysis	Self-organizing map (SOM)	Accuracy = 97.1%

## REFERENCES

- [1] H. Zarafshan, M. R. Mohammadi, S. A. Motevalian, F. Abolhassani, A. Khaleghi, and V. Sharifi, "Autism research in Iran: a scientometric study," *Iranian Journal of Psychiatry and Behavioral Sciences*, vol. 11, no. 2, 2017.
- [2] M. R. Mohammadi et al., "Prevalence of autism and its comorbidities and the relationship with maternal psychopathology: a national population-based study," *Arch Iran Med*, vol. 22, no. 10, 2019.
- [3] M. R. Mohammadi et al., "Prevalence and correlates of psychiatric disorders in a national survey of Iranian children and adolescents," *Iranian journal of psychiatry*, vol. 14, no. 1, p. 1, 2019.
- [4] S. Talepasand et al., "Psychiatric disorders in children and adolescents: prevalence and sociodemographic correlates in Semnan Province in Iran," *Asian journal of psychiatry*, vol. 40, pp. 9-14, 2019.
- [5] H. Zarafshan, M. R. Mohammadi, F. Abolhassani, S. A. Motevalian, and V. Sharifi, "Developing a comprehensive evidence-based service package for toddlers with autism in a low resource setting: Early detection, early intervention, and care coordination," *Iranian Journal of Psychiatry*, vol. 14, no. 2, p. 120, 2019.
- [6] N. Yousefi, H. Dadgar, M. R. Mohammadi, N. Jalilevand, M. R. Keyhani, and A. Mehri, "The validity and reliability of Autism Behavior Checklist in Iran," *Iranian journal of Psychiatry*, vol. 10, no. 3, p. 144, 2015.
- [7] A. Khaleghi, H. Zarafshan, S. R. Vand, and M. R. Mohammadi, "Effects of non-invasive neurostimulation on autism spectrum disorder: a systematic review," *Clinical Psychopharmacology and Neuroscience*, vol. 18, no. 4, p. 527, 2020.
- [8] A. Ali, F. F. Negin, F. F. Bremond, and S. Thümmel, "Video-based behavior understanding of children for objective diagnosis of autism," in *VISAPP 2022-17th International Conference on Computer Vision Theory and Applications*, 2022.
- [9] H. V. Ratajczak, "Theoretical aspects of autism: biomarkers—a review," *Journal of immunotoxicology*, vol. 8, no. 1, pp. 80-94, 2011.
- [10] D.-Y. Song, S. Y. Kim, G. Bong, J. M. Kim, and H. J. Yoo, "The use of artificial intelligence in screening and diagnosis of autism spectrum disorder: a literature review," *Journal of the Korean Academy of Child and Adolescent Psychiatry*, vol. 30, no. 4, p. 145, 2019.
- [11] K.-F. Kollias, C. K. Syriopoulou-Delli, P. Sarigiannidis, and G. F. Fragulis, "The contribution of machine learning and eye-tracking technology in autism spectrum disorder research: A systematic review," *Electronics*, vol. 10, no. 23, p. 2982, 2021.
- [12] M. Lai et al., "A machine learning approach for retinal images analysis as an objective screening method for children with autism spectrum disorder," *EClinicalMedicine*, vol. 28, p. 100588, 2020.
- [13] Z. Zhao et al., "Applying machine learning to identify autism with restricted kinematic features," *IEEE Access*, vol. 7, pp. 157614-157622, 2019.
- [14] G. Brihadiswaran, D. Haputhanthri, S. Gunathilaka, D. Meedeniya, and S. Jayarathna, "EEG-based processing and classification methodologies for autism spectrum disorder: A review," *Journal of Computer Science*, vol. 15, no. 8, 2019.
- [15] A. Khaleghi, M. R. Mohammadi, K. Shahi, and A. Motie Nasrabadi, "A neuronal population model based on cellular automata to simulate the electrical waves of the brain," *Waves in Random and Complex Media*, pp. 1-20, 2021.
- [16] A. Khaleghi, M. R. Mohammadi, K. Shahi, and A. M. Nasrabadi, "Possible Neuropathological Mechanisms Underlying the Increased Complexity of Brain Electrical Activity in Schizophrenia: A Computational Study," *Iranian Journal of Psychiatry*, pp. 1-7, 2023.
- [17] A. Khaleghi, A. Sheikhan, M. R. Mohammadi, and A. M. Nasrabadi, "Evaluation of cerebral cortex function in clients with bipolar mood disorder I (BMD I) compared with BMD II using QEEG analysis," *Iranian Journal of Psychiatry*, vol. 10, no. 2, p. 93, 2015.
- [18] A. Khaleghi, H. Zarafshan, and M. R. Mohammadi, "Visual and auditory steady-state responses in attention-deficit/hyperactivity disorder," *European archives of psychiatry and clinical neuroscience*, vol. 269, pp. 645-655, 2019.
- [19] M. Moeini, A. Khaleghi, N. Amiri, and Z. Niknam, "Quantitative electroencephalogram (QEEG) spectrum analysis of patients with schizoaffective disorder compared to normal subjects," *Iranian Journal of Psychiatry*, vol. 9, no. 4, p. 216, 2014.
- [20] M. Moeini, A. Khaleghi, and M. R. Mohammadi, "Characteristics of alpha band frequency in adolescents with bipolar II disorder: a resting-state QEEG study," *Iranian journal of psychiatry*, vol. 10, no. 1, p. 8, 2015.
- [21] M. Moeini, A. Khaleghi, M. R. Mohammadi, H. Zarafshan, R. L. Fazio, and H. Majidi, "Cortical alpha activity in schizoaffective patients," *Iranian Journal of Psychiatry*, vol. 12, no. 1, p. 1, 2017.
- [22] M. R. Mohammadi, A. Khaleghi, A. M. Nasrabadi, S. Raffieivand, M. Begol, and H. Zarafshan, "EEG classification of ADHD and normal children using non-linear features and neural network," *Biomedical Engineering Letters*, vol. 6, pp. 66-73, 2016.
- [23] W. Xiao, G. Manyi, and A. Khaleghi, "Deficits in auditory and visual steady-state responses in adolescents with bipolar disorder," *Journal of Psychiatric Research*, vol. 151, pp. 368-376, 2022.
- [24] A. Afzali, A. Khaleghi, B. Hatef, R. Akbari Movahed, and G. Pirzad Jahromi, "Automated major depressive disorder diagnosis using a dual-input deep learning model and image generation from EEG signals," *Waves in Random and Complex Media*, pp. 1-16, 2023.
- [25] M. R. Mohammadi and A. Khaleghi, "Transsexualism: A different viewpoint to brain changes," *Clinical Psychopharmacology and Neuroscience*, vol. 16, no. 2, p. 136, 2018.
- [26] W. J. Bosl, H. Tager-Flusberg, and C. A. Nelson, "EEG analytics for early detection of autism spectrum disorder: a data-driven approach," *Scientific reports*, vol. 8, no. 1, pp. 1-20, 2018.
- [27] D. Haputhanthri, G. Brihadiswaran, S. Gunathilaka, D. Meedeniya, Y. Jayawardena, S. Jayarathna, and M. Jaime, "An EEG based channel optimized classification approach for autism spectrum disorder," in *2019 Moratuwa Engineering Research Conference (MERCon)*, 2019: IEEE, pp. 123-128.
- [28] M. Ahmadi, H. Adeli, and A. Adeli, "Fuzzy synchronization likelihood-wavelet methodology for diagnosis of autism spectrum disorder," *Journal of neuroscience methods*, vol. 211, no. 2, pp. 203-209, 2012.
- [29] T.-H. Pham et al., "Autism spectrum disorder diagnostic system using HOS bispectrum with EEG signals," *International journal of environmental research and public health*, vol. 17, no. 3, p. 971, 2020.

- [30] M. Baygin et al., "Automated ASD detection using hybrid deep lightweight features extracted from EEG signals," *Computers in Biology and Medicine*, vol. 134, p. 104548, 2021.
- [31] N. Alotaibi and K. Maharatna, "Classification of autism spectrum disorder from EEG-based functional brain connectivity analysis," *Neural Computation*, vol. 33, no. 7, pp. 1914-1941, 2021.
- [32] M. Radhakrishnan, K. Ramamurthy, K. K. Choudhury, D. Won, and T. A. Manoharan, "Performance Analysis of Deep Learning Models for Detection of Autism Spectrum Disorder from EEG Signals," *Traitement du Signal*, vol. 38, no. 3, 2021.
- [33] A. Khaleghi, P. M. Birgani, M. F. Fooladi, and M. R. Mohammadi, "Applicable features of electroencephalogram for ADHD diagnosis," *Research on Biomedical Engineering*, vol. 36, pp. 1-11, 2020.
- [34] A. Khaleghi, M. R. Mohammadi, M. Moeini, H. Zarafshan, and M. Fadaei Fooladi, "Abnormalities of alpha activity in frontocentral region of the brain as a biomarker to diagnose adolescents with bipolar disorder," *Clinical EEG and neuroscience*, vol. 50, no. 5, pp. 311-318, 2019.
- [35] A. Khaleghi, A. Sheikhani, M. R. Mohammadi, A. M. Nasrabadi, S. R. Vand, H. Zarafshan, and M. Moeini, "EEG classification of adolescents with type I and type II of bipolar disorder," *Australasian physical & engineering sciences in medicine*, vol. 38, pp. 551-559, 2015.
- [36] H. Zarafshan, A. Khaleghi, M. R. Mohammadi, M. Moeini, and N. Malmir, "Electroencephalogram complexity analysis in children with attention-deficit/hyperactivity disorder during a visual cognitive task," *Journal of clinical and experimental neuropsychology*, vol. 38, no. 3, pp. 361-369, 2016.
- [37] V. Srinivasan, C. Eswaran, and N. Sriraam, "Approximate entropy-based epileptic EEG detection using artificial neural networks," *IEEE Transactions on information Technology in Biomedicine*, vol. 11, no. 3, pp. 288-295, 2007.
- [38] A. Rizal and S. Hadiyoso, "Sample entropy on multidistance signal level difference for epileptic EEG classification," *The Scientific World Journal*, vol. 2018, 2018.
- [39] A. Pavlov, A. Runnova, V. Maksimenko, O. Pavlova, D. Grishina, and A. Hramov, "Detrended fluctuation analysis of EEG patterns associated with real and imaginary arm movements," *Physica A: Statistical Mechanics and its Applications*, vol. 509, pp. 777-782, 2018.
- [40] S. Hatamikia and A. M. Nasrabadi, "Recognition of emotional states in response to audio-visual inductions based on nonlinear analysis and self-organisation map classification," *International Journal of Medical Engineering and Informatics*, vol. 9, no. 2, pp. 110-133, 2017.
- [41] N. J. Minshew and T. A. Keller, "The nature of brain dysfunction in autism: functional brain imaging studies," *Current opinion in neurology*, vol. 23, no. 2, pp. 124-130, 2010.
- [42] A. M. Kopec, M. R. Fiorentino, and S. D. Bilbo, "Gut-immune-brain dysfunction in autism: importance of sex," *Brain research*, vol. 1693, pp. 214-217, 2018.
- [43] F. C. Peck, L. J. Gabard-Durnam, C. L. Wilkinson, W. Bosl, H. Tager-Flusberg, and C. A. Nelson, "Prediction of autism spectrum disorder diagnosis using nonlinear measures of language-related EEG at 6 and 12 months," *Journal of neurodevelopmental disorders*, vol. 13, no. 1, pp. 1-13, 2021.
- [44] W. J. Bosl, T. Loddenkemper, and C. A. Nelson, "Nonlinear EEG biomarker profiles for autism and absence epilepsy," *Neuropsychiatric Electrophysiology*, vol. 3, no. 1, pp. 1-22, 2017.
- [45] T. Heunis, C. Aldrich, J. Peters, S. Jeste, M. Sahin, C. Scheffer, and P. De Vries, "Recurrence quantification analysis of resting state EEG signals in autism spectrum disorder—a systematic methodological exploration of technical and demographic confounders in the search for biomarkers," *BMC medicine*, vol. 16, pp. 1-17, 2018.
- [46] S. L. Oh et al., "A novel automated autism spectrum disorder detection system," *Complex & Intelligent Systems*, vol. 7, no. 5, pp. 2399-2413, 2021.
- [47] L. O.-S. C. T. Keh, A. M. A. Chupungco, and J. P. Esguerra, "Nonlinear time series analysis of electroencephalogram tracings of children with autism," *International Journal of Bifurcation and Chaos*, vol. 22, no. 03, p. 1250044, 2012.
- [48] M. Ahmadlou, H. Adeli, and A. Adeli, "Fractality and a wavelet-chaos-neural network methodology for EEG-based diagnosis of autistic spectrum disorder," *Journal of Clinical Neurophysiology*, vol. 27, no. 5, pp. 328-333, 2010.
- [49] W. Bosl, A. Tierney, H. Tager-Flusberg, and C. Nelson, "EEG complexity as a biomarker for autism spectrum disorder risk," *BMC medicine*, vol. 9, no. 1, pp. 1-16, 2011.
- [50] A. Sheikhani, H. Behnam, M. R. Mohammadi, M. Noroozian, and M. Mohammadi, "Detection of abnormalities for diagnosing of children with autism disorders using of quantitative electroencephalography analysis," *Journal of medical systems*, vol. 36, pp. 957-963, 2012.
- [51] W. Jamal, S. Das, I.-A. Oprescu, K. Maharatna, F. Apicella, and F. Sicca, "Classification of autism spectrum disorder using supervised learning of brain connectivity measures extracted from synchrostates," *Journal of neural engineering*, vol. 11, no. 4, p. 046019, 2014.
- [52] J. Eldridge, A. E. Lane, M. Belkin, and S. Dennis, "Robust features for the automatic identification of autism spectrum disorder in children," *Journal of neurodevelopmental disorders*, vol. 6, no. 1, pp. 1-12, 2014.
- [53] J. Kang, T. Zhou, J. Han, and X. Li, "EEG-based multi-feature fusion assessment for autism," *Journal of Clinical Neuroscience*, vol. 56, pp. 101-107, 2018.

# A Composite Noise Removal Network Based on Multi-domain Adaptation

Fan Bai<sup>1</sup>, Pengfei Li<sup>2\*</sup>, Haoyang Sun<sup>3</sup>, Hui Zhang<sup>4</sup>

School of Equipment Engineering, Shenyang Ligong University, Shenyang, China<sup>1,4</sup>  
Science and Technology on Electromechanical Dynamic Control Laboratory<sup>2</sup>

School of Mechanical and Electrical Engineering, Beijing Institute of Technology, Xi'an Beijing, China<sup>2</sup>  
School of Mechanical and Electrical Engineering, Beijing Institute of Technology Beijing, China<sup>3</sup>

**Abstract**—Addressing the limitation of conventional single-scene image denoising algorithms in filtering mixed environmental disturbances, and recognizing the drawbacks of cascaded image enhancement algorithms, which have poor real-time performance and high computational demands, The composite weather adaptive denoising network (CWADN) is proposed. A Cascade Hourglass Feature Extraction Network is constructed with a visual attention mechanism to extract characteristics of rain, fog, and low-light noise from authentic natural images. These features are then transferred from their original real distribution domain to a synthetic distribution domain using a deep residual convolutional neural network. The generator and style encoder of the adversarial network work together to adaptively remove the transferred noise through a combination of supervised and unsupervised training, this approach achieves adaptive denoising capabilities tailored to complex natural environmental noise. Experimental results demonstrate that the proposed denoising network yields a high signal-to-noise ratio while maintaining excellent image fidelity. It effectively prevents image distortion, particularly in critical target areas. Additionally, it adapts to various types of mixed noise, making it a valuable tool for preprocessing images in advanced machine vision algorithms such as target recognition and tracking.

**Keywords**—Image denoising; domain adaptation; generative adversarial network; autoencoder

## I. INTRODUCTION

Haze, rain, and low illumination are the three types of natural noises that have the greatest impact on the detection accuracy of machine vision. These noises will destroy the optical information in the original image through global blurring, superimposed noise, and information desalination, bringing a great challenge to all-weather target detection tasks.[1], [2]. Therefore, the denoising methods for the above three natural noises have become the key research directions of domestic and foreign scholars in the field of image denoising. Among them, the study [3] directly learns and estimates the mapping function between the noisy image and its noise-free counterpart and cooperates with the bilateral rectified linear unit (BReLU) to reduce the search space and improve the convergence, to realize the end-to-end training and interference process of the dehazing network. Based on the transmittance parameters and atmospheric light of the scattering model, the research in [4] directly learned the residual information between the haze image and the haze-free image by using

smooth dilated convolution and threshold fusion sub-networks to realize image dehazing. The study in [5] pass a binary rain mask to the multi-task network for learning, and the negative rain layer generated by iteration is compared with the input, which reduces the effect of rain noise on the original image. Based on the reverse stacking denoising strategy, the study in [6] use a dataset marked with the rain size to train a multi-task network and realizes image denoising through the obtained rain noise features, [7] use Retinex theory, the reflectance image under ideal illumination is multiplied by the noisy low-illumination image, and the low-illumination noise is directly removed by guided filtering, which solves the problem that the traditional low-illumination denoising algorithm over-enhances or under-enhances some areas. Enhanced question in study [8] constructed an unsupervised network EnlightenGAN, which can be trained in a large number of imprecisely matched images to establish a mapping relationship, which overcomes the problem of low-illumination noise denoising accuracy when the dataset is insufficient. The study in [9] designed a dual-branch unit endowed with physics-aware, complemented by a course learning contrast regularization approach. This research underscores the significance of fine-tuning various negative samples within the contrast regularization process. These insights offer valuable concepts for leveraging multimodal contrastive regularization techniques to enhance image quality.

However, because the environmental noise in nature appears in the form of mixed accompaniment, that is the three kinds of noises of haze, rain and low illumination may be generated at the same time in different weather and will be mapped on the original image in the form of mixing in any proportion. The interference of noise on image information will become more complicated. The above algorithms all adopt the directional denoising strategy, and it is difficult to achieve an optical result when it comes to mixed noise in real scenes. Therefore, researchers gradually focus their research on the field of adaptive denoising that is more in line with actual needs and can integrate various physical models of weather noise. The research in [10] applied the strategy of Neural Architecture Search in reinforcement learning to image restoration to generate the most suitable denoising network structure, and at the vector level, Denoising the output results of Encoders), giving the algorithm the ability of adaptive filtering of compound noise; The study [11] introduced an attention mechanism (Spatial Attention Mechanism, SAM) and



cross-multi-stage feature fusion in the encoding and decoding process of the network. The mechanism (Cross-stage Feature Fusion, CSFF) avoids the loss of target feature information before and after denoising. It takes into account the functions of efficient denoising and target information transfer. Although these composite noise-denoising deep neural networks have good image processing capabilities, their overly complex structures lead to high demands on computing resources. The training difficulty and convergence speed are not ideal. At the same time, a large number of supervised learning links make this kind of network must be supported by abundant real noise datasets to obtain better training results. When the real noisy images of the actual scene are difficult to obtain, the noise reduction accuracy of this kind of network will be greatly improved. These problems will limit the versatility of denoising algorithms in real environments. Providing all-weather adaptive denoising capabilities for platforms such as space vehicles and ground-based photodetectors is difficult.

In order to solve the problems of the above algorithms and improve the adaptability and all-weather computing efficiency of the image denoising algorithm in the denoising task for complex natural environments, this paper proposes a denoising method for the free mixed environment noise of rain, haze and low illumination. Noise neural network, the innovations of this network are: (1) An end-to-end image denoising network based on domain transfer is proposed, which realizes rain, haze, low illumination and three kinds of mixed noise images under a single structure. (2) Integrate multi-stage autoencoder structure and multi-domain transfer strategy to achieve directional separation and targeted denoising of an unknown proportion of mixed noise. (3) Based on domain adaptive generative confrontation module, effectively reduce the difference between noisy synthetic data and real noisy data is eliminated, and the traditional denoising algorithm training process is free

from the dependence on a large number of real noisy data sets. Based on the above methods and characteristics, the denoising network proposed in this paper achieves a high signal-to-noise ratio and image structure consistency in multi-type mixed noise filtering tasks and achieves high-quality denoising and information restoration for complex natural environment noise.

## II. PROPOSED METHODS

This paper draws extensive inspiration from the multi-level architecture of MPRNet [12], which strikes a balance between preserving local and global information. It introduces the concept of projecting any natural environmental image into multiple modes, followed by individual processing and subsequent integration. The composite weather adaptive denoising network (CWADN) designed in this paper is composed of a separation module, a denoising module and a conversion module, and its network structure is shown in Fig. 1.

CWADN comprises three integral components: the Multi-stage Progressive Separation Network (MPSN), the Multi-domain Translation Noise Network (MTDN), and the Domain Adaptation Translation Network (DATN). In the denoising process, CWADN initially takes images containing complex real-world composite noise as input into the multi-level autoencoder of MPSN. Subsequently, MTDN leverages the noise distribution within the image space as a feature and transfers the three distinct noise images to the synthetic domain for generation. Finally, DATN restores these three transformed images, consolidates the acquired results, and accomplishes the denoising task. Through these methodologies, CWADN achieves adaptive and precise denoising, as well as the restoration of target image information for original images captured in real-world scenarios featuring natural compound noise.

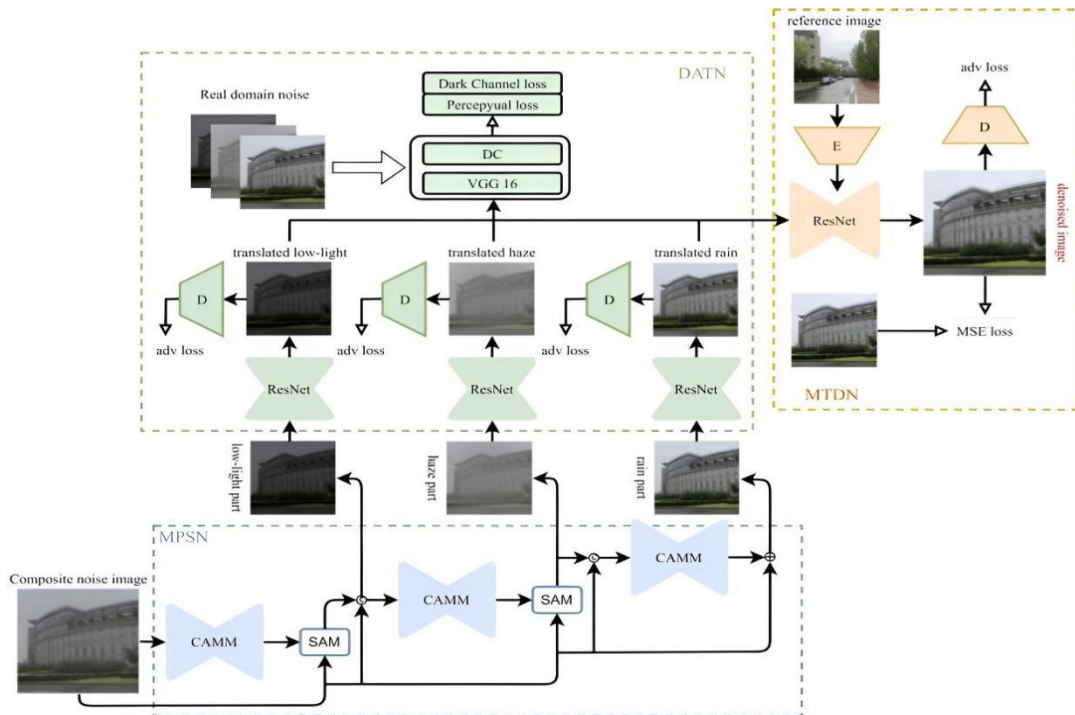


Fig. 1. The framework of our proposed network.

### III. NETWORK DETAILS

#### A. Multi-Stage Progressive Separation Network

For any noisy image shot in a natural environment, it can be regarded as the result of the original noise-free image.  $I_{ori}$  is affected by three kinds of noise: rain, haze and low illumination, and different types of noise have different effects on  $I_{ori}$ . The image quality degradation process can be expressed as:

$$I_n = \lambda_{rain} N_{rain}(I_{ori}) \odot \lambda_{haze} N_{haze}(I_{ori}) \odot \lambda_{dark} N_{dark}(I_{ori}) \quad (1)$$

Where  $\odot$  is the mixing operation of noise,  $N_i$  and  $\lambda_i$  are the noise degradation sub-function and the noise component weight coefficient under the conditions of rain, haze, and low illumination ( $i = \{rain, haze, dark\}$ ), respectively. When there is no certain kind of noise, the trade-off  $\lambda_i$  equals to 0, and when the image is an ideal noise-free image, all  $\lambda$  are 0.

Due to the diversity and unpredictability of the combination of each proportional coefficient  $\lambda_i$  of the composite noise in the natural environment, it is difficult for the conventional denoising network to fit the image quality degradation function with a certain coefficient, and it is impossible to learn the mapping relationship between  $I_n$  and  $I_{ori}$  or complete image restoration. To solve this problem, we design a multi-stage progressive separation network (MPSN) to separate the composite noise and sequentially extract the low-illumination noise component  $N_{dark}$ , the haze noise component  $N_{haze}$  and the rain noise component  $N_{rain}$ , and the different noise components

are limited to their own domain according to their characteristics. MPSN is composed of three cascaded Channel Attention Autoencoder Module (CAAM), and its network structure is shown in Fig. 2.

In the autoencoder network based on the full convolution layer, although the continuous convolution operation can enrich the semantic information of the feature map, it also causes the gradual loss of the texture information in the original image, which makes the deconvolution operation unable to correct the decoding process and accurately restored details of the image. Therefore, we add a Channel Attention Block (CAB) [13] to all convolution and deconvolution operations in all CAAM to reduce information loss in key areas of interest. The input feature map group of CAB is denoted as  $X = [x_1, x_2, \dots, x_n]$ , where  $x_n$  is the feature map of the  $n$  channel with size  $H \times W$ . The frequency information of features is included in  $x_n$ , and its high-frequency features can better represent the edge and detail information in the image. Therefore, the global average pooling obtains the global feature frequency  $z_n$  by scaling the size of  $x_n$  to  $1 \times 1$ . Then, in order to extract the channel feature of  $z_n$  and obtain the weight coefficient through the activation operation, CAB will adjust the weight of  $x_n$  to obtain the final feature  $x_n^*$  with the attention mechanism. This process can be expressed as:

$$\begin{cases} z_n = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W x_n(i, j) \\ x_n^* = x_n \otimes \sigma(\phi(z_n)) \end{cases} \quad (2)$$

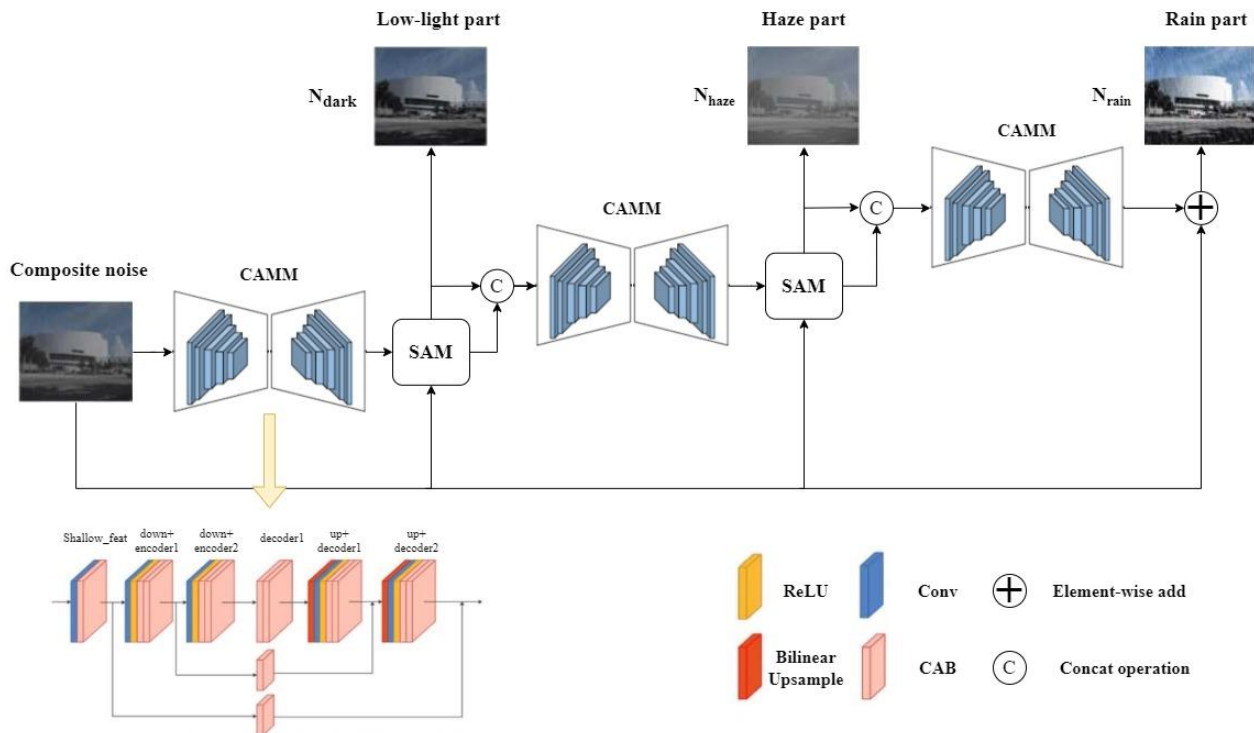


Fig. 2. Channel attention autoencoder Module for our network.

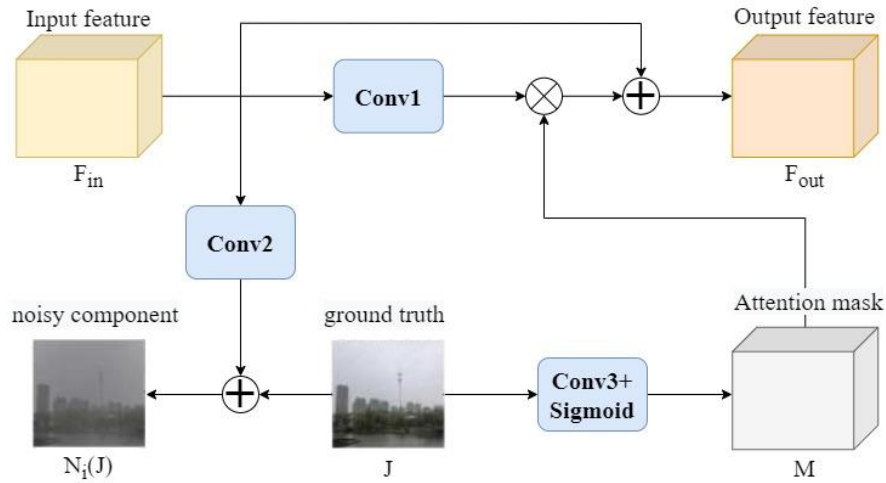


Fig. 3. Spatial attention module for our network.

Where  $i$  and  $j$  represent a pixel's horizontal and vertical position coordinates in the current image, respectively;  $\sigma$  and  $\phi$  represent the sigmoid activation function and the channel feature extraction process, respectively, and  $\otimes$  is the corresponding multiplication on pixel-wise. Through the above attention mechanism, MPSN can assign higher weights to feature channels with high-frequency information in the reconstruction process so as to preserve the texture details of the region of interest.

In addition, we introduce a Spatial Attention Module [14] between each two CAMM to enhance the transfer and fusion of information between different stages. As shown in Fig. 3, SAM obtains the noise component  $N_i$  by performing channel dimension reduction on the input feature map  $F_{in}$  and adding it pixel-by-pixel with the noise-free image  $I_{ori}$ . At the same time, SAM extracts features from  $I_{ori}$  and activates it to obtain the attention mask  $M$  and then linearly changes  $F_{in}$  to obtain the output feature map  $F_{out}$  as the input of the next stage. The process is expressed as:

$$\begin{cases} N_i(I_{ori}) = W_2(F_{in}) + I_{ori} \\ M = \text{sigmoid}(W_3(I_{ori})) \\ F_{out} = F_{in} + W_1(F_{in}) \times M \end{cases} \quad (3)$$

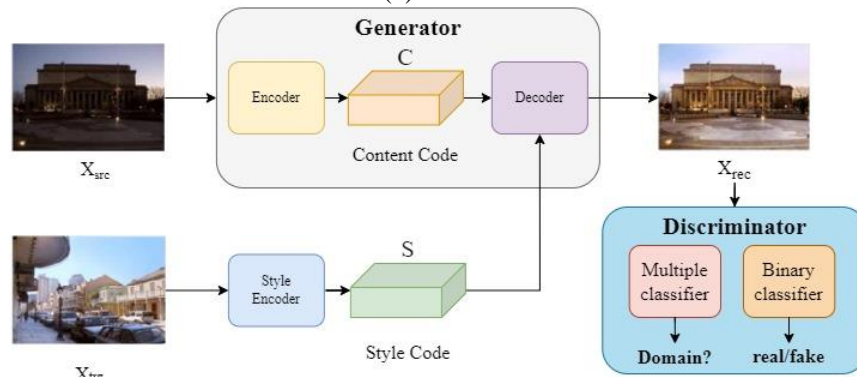


Fig. 4. The framework of multi-domain translation network.

where, the  $W_1$ ,  $W_2$ , and  $W_3$  represent three convolution operations, respectively. After passing through the SAM module,  $F_{out}$  contains ground truth information, which can improve the texture feature representation ability of the subsequent separated images. Due to the constraint of the attention mask, the transmission of invalid information to the next stage is suppressed. After the original noisy image is separated by MPSN, three images containing only a single type of independent noise can be obtained, which effectively solves the problem that the composite noise image quality degradation function is difficult to fit and limits the scope of the solution space, and facilitates the subsequent denoising network.

### B. Multi-Domain Translation Denoise Network

At present, the image restoration methods using image translation mainly regard the noisy image and the noise-free image as two independent domains and use the generator to learn the mapping relationship between them to realize image denoising. Still, this method can only remove a single type of noise. In order to realize the adaptive removal of multiple types of composite noise under a single model, inspired by StarGANv2 [15], we construct a multi-domain translation network (MTDN), including generator  $G$ , style encoder  $E$  and discriminator  $D$ , three main parts which are shown in Fig. 4.

We take the samples of noisy images, which include rain, haze and low illumination as the source domain  $X_{src}$ , that is,  $X_{src} = \{N_{haze}, N_{rain}, N_{dark}\}$ , and the noise-free images as the target domain  $X_{trg} = I_{org}$ , by learning the mapping function between  $X_{src}$  and  $X_{trg}$  to realize the removal of many different types of noises. E extracts features from the target domain image  $x_{trg}$  ( $x_{trg} \in X_{trg}$ ), and obtains the style code  $s$  containing the high-dimensional feature information of the noise-free image. G extracts the high-dimensional semantic information  $c$  of the original noise image  $x_{src}$  ( $x_{src} \in X_{src}$ ) through internal coding, integrates  $c$  and  $s$  in the decoder, and decodes the reconstructed image of  $x_{rec}$ . As a multi-task discriminator,  $D$  is composed of a binary classifier and a multi-classifier. The multi-classifier is used to determine whether the  $x_{rec}$  has the characteristics of a noise-free image, that is, whether the denoising process is complete; the binary classifier evaluates the image quality of

the  $x_{rec}$ , that is, whether the reconstructed image has higher restoration. Through the above process, MTDN can achieve directional denoising for various types of noises and effectively retain the original image's texture, details and other information.

### C. Domain adaptation translation network

Due to the inability to obtain a large number of real noisy-clean paired datasets, both MPSN and MTDN can only be trained on synthetic datasets. However, the spatial distribution and texture features of environmental noise in the synthetic dataset are different from those in the real dataset. This makes the model trained on the synthetic dataset generalize well to the real-world samples and reduces the denoising performance. In order to solve the above problems, we use DATN to improve the adaptability of the model in different datasets.

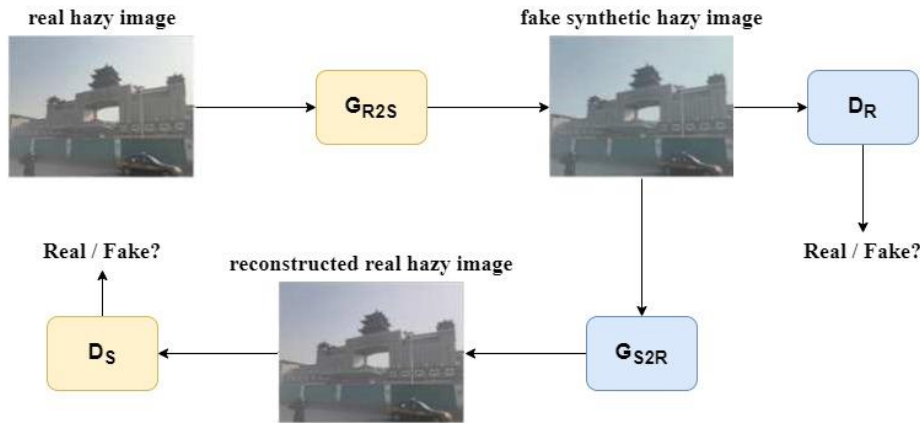


Fig. 5. The framework of domain adaptation translation network.

The structure of DATN is shown in Fig. 5, which is composed of three translation modules with the same structure, corresponding to three different types of noise, respectively. DATN takes CycleGAN [16] as the core framework, which contains two generators,  $G_{S2R}$ ,  $G_{R2S}$  and two discriminators,  $D_R$  and  $D_S$ . The generator  $G_{S2R}$  takes the synthetic noise  $N_{syn}$  as input and uses the spatial distribution characteristics of noise in the real noise image  $N_{rea}$  as the learning object to reconstruct  $N_{syn}$ . The discriminator  $D_R$  compares the difference between the reconstructed image and  $N_{rea}$  and uses it as an indicator to constrain  $G_{S2R}$ . During training,  $G_{S2R}$  improves its domain adaptation ability by continuously increasing the similarity between the generated image and  $N_{rea}$ . Similarly, the same process will be used for training and calculation for  $G_{R2S}$  and  $D_S$ . Therefore, DATN can realize the mutual translation between real and synthetic noise images through a large number of unpaired datasets, thereby improving the generalization performance of CWADN in real noise samples.

### D. Loss function

In order to obtain the best training effect of each functional module and improve the convergence ability and subsequent calculation accuracy of network denoising learning, this paper analyzes the loss of each sub-network in the proposed CWADN according to the task types and structural characteristics of different functional modules. The function has been specially designed.

### Image separation Losses

We expect MPSN to separate three different kinds of noise  $N_i$  from composite noise. We train the MPSN in a supervised manner due to the synthetic training samples. Firstly, we calculate the mean square loss between MPSN result  $N_i^{out}$  and ground truth  $N_i^{gt}$

$$L_{mse}^{MPSN} = \left\| N_i^{out} - N_i^{gt} \right\|_2^2 \quad (4)$$

Where  $i = \{haze, rain, dark\}$  represents the three noise conditions under haze, rain and low illumination, respectively. In order to prevent the noise separation process from destroying the information of non-noise areas in the original image, we use structural loss  $L_{ssim}^{MPSN}$  to constrain the distortion between  $N_i^{out}$  and  $N_i^{gt}$ , and judge the image quality after noise separation.

$$L_{ssim}^{MPSN} = 1 - SSIM(N_i^{out}, N_i^{gt}) \quad (5)$$

Where SSIM is the image similarity calculation rule, which is based on the distribution of image data in the mean, variance and covariance, compared the difference between  $N_i^{out}$  and

$N_i^{gt}$  in lighting, contrast, and image structure, the computing process can be defined as:

$$SSIM(N_i^{out}, N_i^{gt}) = \frac{(2\mu_{out}\mu_{gt} + c_1)}{(\mu_{out}^2 + \mu_{gt}^2 + c_1)} \cdot \frac{(2\sigma_{out,gt} + c_2)}{(\sigma_{out}^2 + \sigma_{gt}^2 + c_2)} \quad (6)$$

Where  $\mu_{out}$  and  $\mu_{gt}$  represent the mean of  $N_i^{out}$  and  $N_i^{gt}$ ,  $\sigma_{out}^2$  and  $\sigma_{gt}^2$  represent the variance of  $N_i^{out}$  and  $N_i^{gt}$ ,  $\sigma_{out,gt}$  is the covariance of  $N_i^{out}$  and  $N_i^{gt}$ .  $c_1$  and  $c_2$  are tiny decimals,  $L$  is the dynamic range of pixel values. When the SSIM value equals to 1, it represents that the MPSN has strong image noise separation and quality restoration capabilities. Therefore, the total loss of MPSN can be expressed as follows:

$$L^{MSPN} = \lambda_{mse} L_{mse}^{MSPN} + \lambda_{ssim} L_{ssim}^{MSPN} \quad (7)$$

Where  $\lambda_{mse}$  and  $\lambda_{ssim}$  represent the trade-off of  $L_{mse}^{MSPN}$  and  $L_{ssim}^{MSPN}$  respectively.

#### Image denoising Losses

For MTDN, unsupervised training can enable the network to achieve denoising by reconstructing images. Still, the reconstructed images lose texture details, affecting the accuracy of subsequent target recognition, tracking and other advanced machine vision [17]. To solve this problem, we combine supervised and unsupervised learning, adopting a semi-supervised learning method to train MTDN. The unsupervised training process uses a generative adversarial loss. The style encoder  $E_{dn}$  maps the target domain image  $x_{trg}$  to the corresponding style code  $s = E_{dn}(x_{trg})$ , and the generator  $G_{dn}$  integrates the original domain images  $x_{src}$  and  $s$ . After reconstruction and denoising, the image  $G_{dn}(x_{src}, s)$  is combined with the discriminator  $G_{dn}$  to get the adversarial loss of the denoising network, which is defined as:

$$L_{adv}^{MTDN} = \mathbb{E}_{x_{src}} [\log(D_{dn}(x_{src}))] + \mathbb{E}_{x_{src}, x_{trg}} [\log(1 - D_{dn}(G_{dn}(x_{src}, s)))] \quad (8)$$

MTDN realizes the adaptive removal of various types of noise by learning the mapping function between  $x_{src}$  and  $x_{trg}$ . However, when only introduce an adversarial loss in the training process; it will cause the generator to confuse the discriminator to get a higher score, which will reduce the problem of the diversity of generated images, called mode collapse. In order to solve this problem, we add the cycle consistency loss; after translating image  $G_{dn}$  to the source domain again, the result should be consistent with source

domain image  $x_{src}$ , that is  $x_{src} = G_{dn}(G_{dn}(x_{src}, s), s^*)$ , the above process can be expressed as:

$$L_{cyc}^{MTDN} = \mathbb{E}_{x_{src}, x_{trg}} [\|x_{src} - G_{dn}(G_{dn}(x_{src}, s), s^*)\|_1] \quad (9)$$

Where  $s^*$  is the style code of the source domain image, that is,  $s^* = E_{dn}(x_{src})$ . In addition, in order to make the style code  $s$  better guide the image reconstruction process, we introduce the style reconstruction loss on the basis of the above, which is similar to the cycle consistency loss. The style code  $s$  of the noise-free image is obtained after  $G_{dn}$ ; when it is encoded by  $E_{dn}$  again, the output result should be less different from  $s$ , that is,  $s^* = E_{dn}(G_{dn}(x_{src}, s))$ , the style reconstruction loss is defined as follows:

$$L_{sty}^{MTDN} = \mathbb{E}_{x,z} [\|s - E_{dn}(G_{dn}(x_{src}, s))\|_1] \quad (10)$$

Through the above unsupervised loss, MTDN can recover images affected by three different noises of rain, haze and low illumination, respectively, and can effectively preserve the content information of the images. In order to further retain the details of the denoised image, we introduce a supervised loss based on the unsupervised loss to calculate the mean square error between the denoised image and its noise-free counterpart Iori and preserve the underlying texture information of the image by minimizing the pixel-by-pixel difference between the two.

$$L_{mse}^{MTDN} = \|G_{dn}(x_{src}, E_{dn}(x_{trg})) - I_{ori}\|_2^2 \quad (11)$$

In summary, the overall loss function of MTDN Loss can be expressed as follows:

$$L^{MTDN} = L_{adv}^{MTDN} + \lambda_{sty} L_{sty}^{MTDN} + \lambda_{cyc} L_{cyc}^{MTDN} + \lambda_{mse} L_{mse}^{MTDN} \quad (12)$$

Where,  $\lambda_{sty}$ ,  $\lambda_{cyc}$ , and  $\lambda_{mse}$  are trade-off weights. During the training process, MTDN minimizes loss to realize the targeted removal of rain, haze and low illumination on the basis of retaining the target texture information.

#### E. Image translation Losses

Noise  $N_i$  can be divided into real noise  $N_{rea}^i$  and synthetic noise  $N_{syn}^i$  that is  $N_i = \{N_{rea}^i, N_{syn}^i\}$ . The training datasets of MPSN and MTDN are all synthetic samples, due to the problem of domain shift between different datasets; the removal effect of the model for real noise has decreased, so the loss function of DATN mainly solves the difference between the real noise domain and the synthetic noise domain. DATN contains three parallel transformation sub-networks, each of which is composed of generators GS2R, GR2S and discriminators DS, DR. For the real noise image  $x_{rea}$  ( $x_{rea} \in N_{rea}^i$ ), the  $GR2S(x_{rea})$  transformed by the generator can be closer to the synthetic noise image  $x_{syn}$  ( $x_{syn} \in N_{syn}^i$ ). The adversarial loss can be expressed as:

$$L^{MTDN} = L_{adv}^{MTDN} + \lambda_{sty} L_{sty}^{MTDN} + \lambda_{cyc} L_{cyc}^{MTDN} + \lambda_{mse} L_{mse}^{MTDN} \quad (13)$$

Similarly, the adversarial loss function of  $x_{syn}$  to  $x_{rea}$  translation can be expressed as:

$$L_{adv}^{DATN}(G_{S2R}, D_R) = \mathbb{E}_{x_{rea}} \|\log D_R(x_{rea})\|_1 + \mathbb{E}_{x_{syn}} \|\log(1 - D_R(G_{S2R}(x_{syn})))\|_1 \quad (14)$$

In addition to the same principle as the denoising module, DATN adopts cycle consistency loss to alleviate the mode collapse problem. For the real noisy image  $x_{rea}$ , after sequentially passing through GR2S and GS2R, the result should be close to the input image, that is,  $x_{rea} \approx \text{GR2S}(\text{GS2R}(x_{rea}))$ , then the cycle consistency loss of DATN can be defined as:

$$L_{adv}^{DATN}(G_{S2R}, D_R) = \mathbb{E}_{x_{rea}} \|\log D_R(x_{rea})\|_1 + \mathbb{E}_{x_{syn}} \|\log(1 - D_R(G_{S2R}(x_{syn})))\|_1 \quad (15)$$

In order to further improve the details of the generated image and enhance the texture feature information of the image, we introduce the perceptual loss to measure the distance of the image before and after transformation in the perceptual feature space; it will not only be limited to the pixel space and the feature extraction will be carried out on the two through the convolutional neural network, but the transformed image will be constrained from the high-dimensional space to make it more stylistically the target domain image.

$$L_{pect}^{DATN} = \mathbb{E}_{x_{rea}, x_{syn}} \frac{1}{CHW} [\|\phi(G_{R2S}(x_{rea})) - \phi(x_{syn})\|_1 + \|\phi(G_{S2R}(x_{syn})) - \phi(x_{rea})\|_1] \quad (16)$$

Where, C, H, and W represent the feature map's channel number, height and width, respectively;  $\phi$  is the feature extraction network. In this paper, VGG16 is used as the network basis, and the high-dimensional perceptual information provided by perceptual loss is used to enhance the high-frequency information of the converted image so as to improve the reconstruction effect of image details.

In addition, the dark channel  $D_c$  (In) of the image with noise can represent the approximate location of the noise distribution in space [16], so this paper proposes the dark channel consistency loss, which limits the dark channel of the image before and after transformation. The L1 loss is used to ensure the consistency of the dark channels of the two so as to strengthen the network's ability to learn the law of noise distribution. Dark channel consistency loss is defined as:

$$L_{dc}^{DATN} = \mathbb{E}_{x_{rea}, x_{syn}} [\|D_c(G_{R2S}(x_{rea})) - D_c(x_{syn})\|_1 + \|D_c(G_{S2R}(x_{syn})) - D_c(x_{rea})\|_1] \quad (17)$$

Where  $D_c$  is the dark channel computing process that can be defined as:

$$D_c(I) = \min_{y \in W(x)} [\min_{c \in \{r, g, b\}} I^c(y)] \quad (18)$$

Where  $x, y$  represents the coordinates of the pixel point;  $I^c(y)$  represents the colour channel of the image  $I$ ;  $W(x)$  represents the sliding window where the pixel point  $x$  is located. Therefore, the overall loss of DATN is defined as:

$$L^{DATN} = L_{adv}^{DATN}(G_{R2S}, D_S) + L_{adv}^{DATN}(G_{S2R}, D_R) + L_{cyc}^{DATN}(G_{R2S}, G_{S2R}) + L_{pect}^{DATN} + L_{dc}^{DATN} \quad (19)$$

In summary, DATN can alleviate the domain shift problem and improve the generalization of models trained on synthetic samples in real scenarios. So far, the design of all loss functions of CWADN has been completed. By training the network to achieve the best convergence of the loss function, the adaptive denoising and information restoration of composite noise in images captured in real scenes can be realized.

#### IV. IMPLEMENTATION

In order to verify the adaptive denoising and image information restoration capabilities of the designed CWADN, this paper uses the noisy dataset to train and test the network. It compares its computing performance with the current state-of-the-art denoising algorithms to prove the feasibility, accuracy and practicality of CWADN.

##### A. Dataset

Due to the lack of various types of natural weather noise image datasets so far, and it is impossible to get a real paired dataset, we design and build a synthetic natural noise dataset, namely Campus. The dataset contains 1009 images taken on Campus as noise-free samples. It augments the data through random cropping, inversion, etc., and according to the atmospheric physical model, by applying rain, haze and lower brightness on the noise-free samples. The method completes the construction of synthetic noise datasets. At the same time, in order to verify the self-adaptive denoising capability of CWADN for composite noise, we randomly generate three decimals with a sum of 1 as the proportional weight in the Campus dataset and mix the three noises of rain, haze and low illumination to generate different noises like Synthetic noise datasets with different types of noise and matching different natural environments. Finally, the obtained 2500 images are paired with the set noise type labels {clean, haze, rain, dark, compound}. Part of the paired data is shown in Fig. 6. In addition, we also randomly selected 1000 samples from the real datasets RESIDE, LOL, and SPA to further verify the computing power of the CWADN conversion module.

##### B. Training Details

The verification platform of this paper is a computer equipped with an NVIDIA GeForce RTX 3080 GPU. The algorithm is written in the Tensorflow framework and uses ADAM as the training optimizer. The batch size is set to 1, and the size of the input image is set to 256×256. The training process adopts two-stage training. In the first stage, MPSN, MTDN and DATN are trained, respectively. We train MPSN firstly, the epoch is set to 100,  $\lambda_{mse}=1.0$ ,  $\lambda_{ssim}=0.5$ , and the learning rate is set to  $2 \times 10^{-4}$ ; Then train the denoising network MTDN, with a total of 20k iterations,  $\lambda_{sty}$ ,  $\lambda_{cyc}$ ,  $\lambda_{mse}$  are 1.0, 1.0, 0.8, respectively, and the learning rate decays linearly from 10

<sup>4</sup> to  $10^{-6}$ ; Finally, the conversion module DATN is trained, and the epoch is set to 150, The learning rate is set to  $2 \times 10^{-4}$ . The second stage uses a small amount of real data sets to fine-tune the network and supervises the results of the current mainstream denoising algorithms as paired real data. In this process, all the models trained in the first stage are imported,

and then the transformation and denoising network parameters are frozen. Only the parameters of the MPSN are updated for 50 epochs. In the second training process, the real paired dataset we used is the results from the mainstream denoise algorithms.

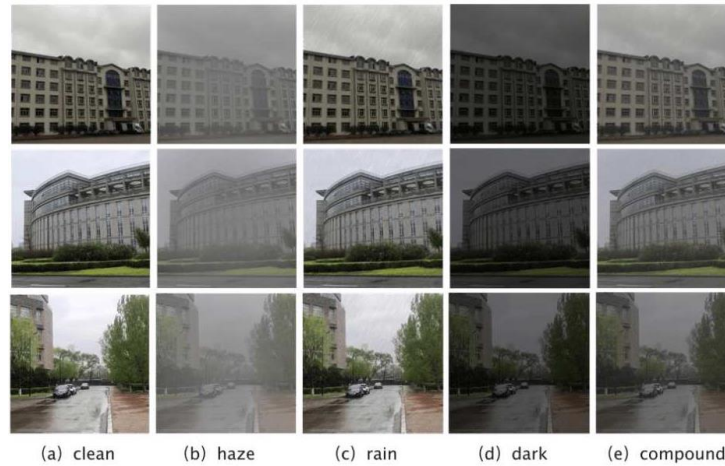


Fig. 6. Samples of the Campus dataset.

## V. RESULTS ANALYSIS

### A. Qualitative Experiment

In this paper, in the real rain, fog, and low illumination scenarios, CWADN and mainstream algorithms in various fields (refer to DCP [18], Cycle GAN, AOD-Net[19] for dehazing capability; reference RESCAN [20] for dehazing capability, PReNet[21], CycleGAN for deraining capability; low illumination enhancement capability is compared with CycleGAN, EnlightenGAN, and Zero-DCE[22]), and the algorithm designed in this paper will be compared from the perspectives of intuitive visual qualitative and image data quantification respectively. The denoising results and computing power are analyzed and judged.

Fig. 7, 8, and 9, respectively, show each algorithm's intuitive visual solution effects in processing images that are degenerated by real haze, rain, and low illumination. Fig. 7 shows the result of the haze removal test of the real sample, from which it can be seen that after DCP calculates the area where the pixel value is close to the atmospheric light value

and the sky area with low contrast, the image after the haze removal will have colour spots and colour shift problems; The image processed by CycleGAN loses more detailed information, and the restored image has low clarity; the image after AOD-Net dehazing is dark as a whole, and the fog noise in some areas is not completely removed; compared with the above algorithm, the saturation and brightness of the image after denoising by CWADN designed in this paper are more natural, and the details of the image are better restored.

Fig. 8 shows the real sample's comparison test of the rain removal effect map. The test image will have rain noise, haze noise and mixed rain and haze noise at the same time. It can be seen from the test results that CycleGAN can filter rain noise and has a certain ability to remove fog noise, but the reconstruction of image texture information is relatively vague; Compared with RESCAN, the rain removal effect of PReNet is significantly improved, but it does not have the ability to remove rain and fog; compared with other algorithms, CWADN has the ability to remove composite noise, it also has a suppressive effect on haze while removing the rain.

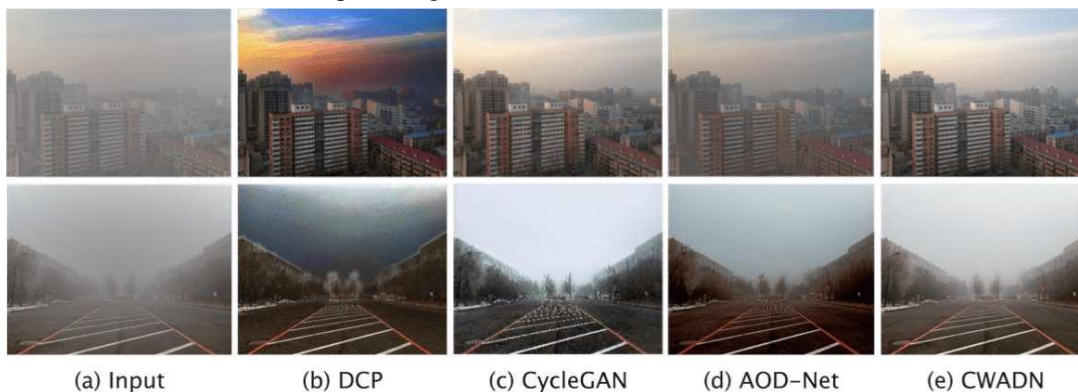


Fig. 7. Comparison of dehazing results on the real-world samples.



Fig. 8. Comparison of deraining results on real-world rain samples.

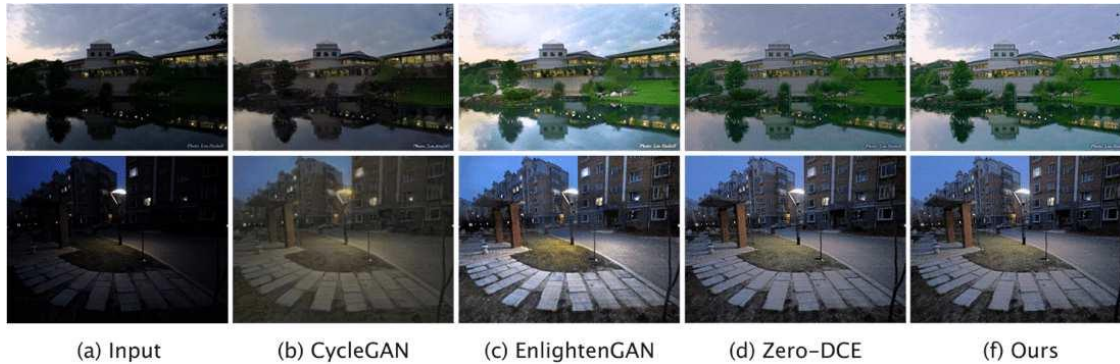


Fig. 9. Comparison of low-light enhancement results on real-world low-light samples.

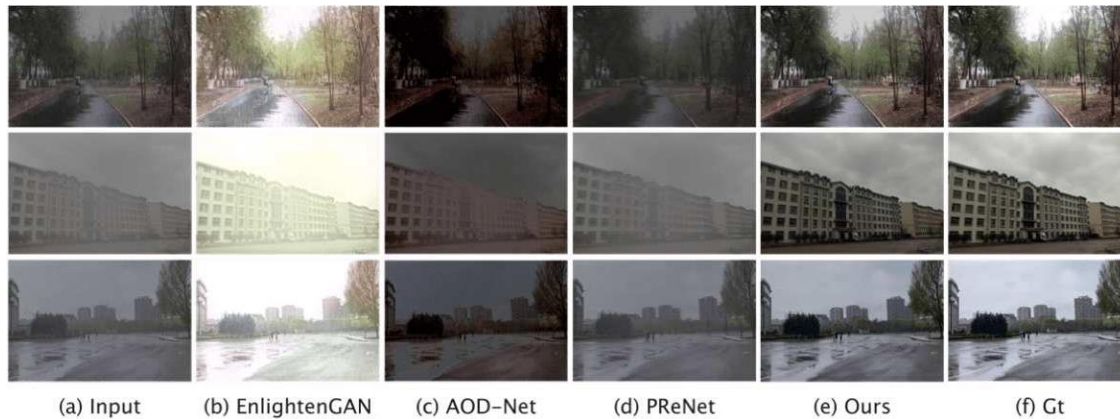


Fig. 10. Comparison of composite noise removal results on Campus dataset.

Fig. 9 shows the low-light enhancement test results under the real sample. It can be seen that the brightness improvement of CycleGAN in some areas is not obvious enough; the saturation of the image enhanced by Zero-DCE is improved, but the overall brightness is low; after EnlightenGAN and CWADN are enhanced. The image achieves good results in both saturation and brightness.

In addition, to further verify the composite noise removal capability of CWADN, we also use the composite noise images in the Campus dataset for testing. In this test, EnlightenGAN, AOD-Net and PReNet were selected for comparison, and the test results are shown in Fig. 10. It can be seen that the image restored by EnligtenGAN is overexposed; the overall brightness of the denoising result of AOD-Net is low, and the

removal effect of dense haze is not ideal; PReNet can effectively remove the rain noise in the image, but there is still a lot of haze noise residual. The above three algorithms cannot effectively remove the image's compound noise. We propose that CWADN is closer to the labelled data image regarding overall visual effect and texture detail retention, which verifies CWADN's denoising ability on composite natural environment noise.

To sum up, through testing on real and synthetic natural noise samples, it can be seen that the CWADN proposed in this paper can remove various types of natural environment noise and various types of mixed noise in any proportion, and its noise adaptive denoising ability can reach even exceeding the effect of some mainstream algorithms, the total subjective



image performance verifies the mixed noise adaptive denoising capability of CWADN.

Intuitive visual qualitative analysis can only judge the denoising and image restoration capabilities of the algorithm from the perspective of macroscopic morphology, and this process is only for the denoising task with the human eye as the observation terminal. For advanced machine vision algorithms such as target recognition and tracking, subsequent calculations need to be performed from the pixel-level data direction after image denoising is completed. In order to further verify the denoising and information restoration capabilities of CWADN at the digital image level, we will further carry out quantitative analysis and evaluation of denoised image data on the basis of the above denoising results.

### B. Quantitative Test

We use Peak Singal to Noise Ratio (PSNR) and Structural Similarity (SSIM) as evaluation indicators to quantitatively verify CWADN and comparison algorithms on three public datasets of SOTS, Rain100H and LOL. The results are shown in Table I, Table II and Table III, respectively. In addition, this paper also tested the effect of each algorithm in removing composite noise on the Campus dataset, and the results are shown in Table IV.

TABLE I. QUANTITATIVE COMPARISON OF DEHAZING RESULTS ON SOTS

Method	PSNR	SSIM
DCP	15.49	0.64
CycleGAN	14.65	0.48
AOD-Net	19.06	0.85
Our method	19.21	0.84

TABLE II. QUANTITATIVE COMPARISON OF DERAINING RESULTS ON RAIN100H

Method	PSNR	SSIM
CycleGAN	24.22	0.77
RESCAN	26.36	0.79
PReNet	26.77	0.86
Our method	26.72	0.86

TABLE III. QUANTITATIVE COMPARISON OF LOW-LIGHT ENHANCEMENT RESULTS ON LOL

Method	PSNR	SSIM
CycleGAN	7.83	0.15
Zero-DCE	14.86	0.59
EnlightenGAN	17.48	0.68
Our method	16.51	0.63

It can be seen from the results in the table that in the dehazing results on the SOTS dataset, the PSNR of CWADN is 19.21dB, which is the best, and the SSIM index is slightly lower than that of AOD-Net; in the rain removal experiment of Rain100H, CWADN and PReNet. The SSIM achieved the best value of 0.86 among all comparison algorithms; in the low-light enhancement comparison results of LOL, the two indicators of EnlightenGAN achieved the best results, and the two indicators of CWADN were slightly lower than those of EnlightenGAN. In the noise-free results, CWADN achieves the

best results in both PSNR and SSIM, with 25.48dB and 0.88, respectively. In Table I, Table II, and Table III, although some indicators of the CWADN designed in this paper are slightly lower than the current mainstream algorithms, the gap between the indicators of CWADN and the mainstream optimal algorithms is small, and the intuitive visual performance and image data calculation are not consistent. No impact. At the same time, Table IV verifies that CWADN has achieved the best performance in the complex noise denoising task, which proves that the algorithm has the ability of all-weather adaptive denoising and is more practical, generalization and denoising than other mainstream algorithms, adaptability.

To sum up, the CWADN proposed in this paper shows good denoising and image restoration capabilities in quantitative experiments, some quantitative detection indicators of denoising exceed the current mainstream algorithms, and the rest performance is on par with the mainstream algorithms. In addition, the experiments on the Campus dataset show that the ability of CWADN to remove mixed noise breaks through the limitation that traditional algorithms are difficult to remove composite noise. Combined with the intuitive visual qualitative analysis results, it is proved that the algorithm in this paper has good noise adaptive filtering and information restoration capabilities in single-type noise directional denoising, multi-type noise denoising and mixed noise denoising tasks, which verifies the performance of the algorithm that is feasibility, practicality and efficiency.

### C. Ablation Study

In order to verify the effectiveness of the components added in MPSN and the loss function added in DATN, a series of ablation experiments are established in this paper. Therefore, a series of ablation experiments are established to test different feature extraction strategies and loss function convergence methods. The effects on MPSN and DATN are, respectively, to verify network performance improvement by the method introduced in this paper.

The effectiveness of the components added in the separation network versus the loss function added in the transformation module.

For MPSN, this paper uses UNet as the basis and gradually increases CAB, SAM and multi-stage strategies. The performance impact of different feature extraction strategies on MPSN is shown in Table IV.

It can be seen from Table IV that after the introduction of CAB, SAM and multi-stage strategies in MPSN, the network performance in PSNR and SSIM is increased by 5.5% (1.555) and 0.43% (0.004), respectively, compared with the standard UNet, which proves that the effectiveness of the feature extraction enhancement strategy introduced in this paper.

TABLE IV. ABLATION RESULTS FOR SEPARATION NETWORK

Method	Multi-stage	SAM	PSNR	SSIM
UNet	-	-	28.649	0.936
UNet+CAB	-	-	28.661	0.821
UNet+CAB	√	-	28.685	0.852
UNet+CAB	√	√	30.206	0.94

TABLE V. ABLATION RESULTS FOR TRANSLATION NETWORK DATN

Loss	PSNR	SSIM
Ladv	13.012	0.463
Ladv +Ldc	13.961	0.496
Ladv +Lpect	23.389	0.923
Ladv +Ldc +Lpect	25.632	0.943

For DATN, based on the unsupervised adversarial loss Ladv, this paper adds the perceptual loss Ldc and the dark channel consistency loss Lpect designed in this paper. The test results are shown in Table V.

It can be seen from Table V that the two added loss functions are helpful to the improvement of network performance, and when the two loss functions of Ldc and Lpect are added at the same time, PSNR and SSIM are nearly doubled compared with the original network. To sum up, by introducing different feature extraction strategies and loss functions into MPSN and DATN, this paper effectively enhances the network denoising performance and convergence state and has a significant improvement effect.

Schemes follow another format (see Fig. 2). If there are multiple panels, they should be listed as (a) a Description of what is contained in the first panel; (b) a Description of what is contained in the second panel. Figures should be placed in the main text near the first time they are cited. A caption on a single line should be centered.

## VI. CONCLUSIONS

In this paper, the composite weather adaptive denoising network (CWADN) is introduced to address the challenge posed by complex mixed weather interference on natural imaging. CWADN leverages cascaded autoencoders with an attention mechanism to effectively separate distinct noise components within mixed noise. Through the adoption of an image translation strategy, a multi-domain denoising network is constructed, enabling adaptive denoising across various noise types. The adaptive domain network structure narrows the gap between real and synthetic noise, thereby enhancing the model's generalization capability in real-world scenarios. This approach combines both supervised and unsupervised techniques during training, yielding excellent results. Experimental findings demonstrate the algorithm's effectiveness in removing single-type and multi-type random mixed noises in real shooting environments while preserving detailed information in key areas of interest. In comparison to traditional algorithms, CWADN exhibits significant improvements in both PSNR and SSIM metrics, highlighting its robust adaptive denoising and image information restoration capabilities.

## VII. LIMITATIONS AND PROSPECTS

While the white-light image enhancement under complex weather conditions is explored in this paper, however, it has been hampered by the limited availability of diverse white-light images captured in varying weather conditions within the same scene. The constrained data volume and scene uniformity have resulted in limited generalization of the training outcomes. Future work aims to address these challenges through extensive data collection efforts, particularly during

nighttime, to curate datasets essential for algorithms in this domain.

Furthermore, considering the difficulties associated with training GAN networks and the inherent blurriness and uncertainties in generated images, the intention is to replace them with state-of-the-art diffusion models in subsequent research. Additionally, the introduction of LLM and video comprehension models into the framework is expected to yield superior image generation results, potentially enhancing the quality of generated images and overall performance.

## ACKNOWLEDGMENT

This work was supported by the project of Science and Technology on Electromechanical Dynamic Control Laboratory, China, No. 6142601220603.

## CONFLICTS OF INTEREST

The authors declared that they have no conflicts of interest.

## REFERENCES

- [1] Y. Hu, Y. Shang, X. Fu, and H. Ding, "A low illumination video enhancement algorithm based on the atmospheric physical model," in 2015 8th International Congress on Image and Signal Processing (CISP), IEEE, 2015, pp. 119–124.
- [2] W. Chen, Z. Jia, J. Yang, and N. K. Kasabov, "Multispectral Image Enhancement Based on the Dark Channel Prior and Bilateral Fractional Differential Model," *Remote Sens (Basel)*, vol. 14, no. 1, p. 233, 2022.
- [3] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "Dehazenet: An end-to-end system for single image haze removal," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5187–5198, 2016.
- [4] D. Chen et al., "Gated context aggregation network for image dehazing and deraining," in 2019 IEEE winter conference on applications of computer vision (WACV), IEEE, 2019, pp. 1375–1383.
- [5] W. Yang, R. T. Tan, J. Feng, J. Liu, Z. Guo, and S. Yan, "Joint rain detection and removal via iterative region dependent multi-task learning," *CoRR*, abs/1609.07769, vol. 2, no. 3, pp. 1–12, 2016.
- [6] H. Zhang and V. M. Patel, "Density-aware single image de-raining using a multi-stream dense network," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 695–704.
- [7] C. Li, J. Guo, F. Porikli, and Y. Pang, "LightNet: A convolutional neural network for weakly illuminated image enhancement," *Pattern Recognit Lett*, vol. 104, pp. 15–22, 2018.
- [8] Y. Jiang et al., "Enlightengan: Deep light enhancement without paired supervision," *IEEE transactions on image processing*, vol. 30, pp. 2340–2349, 2021.
- [9] Yu Zheng et al., "Curricular Contrastive Regularization for Physics-aware Single Image Dehazing" *IEEE Conference on Computer Vision & Pattern Recognition*. IEEE, 2023.
- [10] R. Li, R. T. Tan, and L.-F. Cheong, "All in one bad weather removal using architectural search," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 3175–3185.
- [11] S. W. Zamir et al., "Multi-stage progressive image restoration," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021, pp. 14821–14831.
- [12] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao, "Multi-Stage Progressive Image Restoration." *IEEE Conference on Computer Vision & Pattern Recognition*. IEEE, 2021.
- [13] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in Proceedings of the European conference on computer vision (ECCV), 2018, pp. 286–301.
- [14] S. Waqas Zamir et al., "Multi-stage progressive image restoration," *arXiv e-prints*, p. arXiv-2102, 2021.

- [15] Y. Choi, Y. Uh, J. Yoo, and J.-W. Ha, "Stargan v2: Diverse image synthesis for multiple domains," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 8188–8197.
- [16] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in Proceedings of the IEEE international conference on computer vision, 2017, pp. 2223–2232.
- [17] H. Wang, Z. Yue, Q. Xie, Q. Zhao, Y. Zheng, and D. Meng, "From rain generation to rain removal," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 14791–14801.
- [18] Kaiming He et al, "Single image haze removal using dark channel prior," IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2009.
- [19] Boyi Li et al, "An All-in-One Network for Dehazing and Beyond," arXiv e-prints, p. arXiv:1707.06543,2017.
- [20] Xia Li et al, "Recurrent Squeeze-and-Excitation Context Aggregation Net for Single Image Deraining," arXiv e-prints, p. arXiv:1807.05698,2018.
- [21] Dongwei Ren et al, "Progressive Image Deraining Networks: A Better and Simpler Baseline," IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2019.
- [22] Chunle Guo et al, "Zero-DCE: Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement," IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2020.